JOURNAL OF THEORETICAL BIOLOGY

ACADEMIC PRESS

VOLUME 1, NUMBER 1 JANUARY 1961

London and New York

Journal of Theoretical Biology

Volume 1, 1961: 121*s.* 6*d.* ($17.00)

Private subscriptions only: 86*s.* ($12.00)

Subscription orders should be sent to Academic Press Inc. (London) Ltd., 17 Old Queen Street, London, S.W.1, except those from Canada, Central and South America, and the U.S.A. which should be sent to Academic Press Inc., 111 Fifth Avenue, New York 3, N.Y.

Preface

This journal has been established as a central publishing medium for theoretical work in all fields of biology.

Whereas there is a multiplicity of journals, and probably an excess of publishing capacity, for many types of experimental work, it has frequently been difficult for an author to find an acceptable medium for theoretical papers. This journal has arisen in response to a wide-spread demand for a journal that would specialize in theoretical work.

Manuscripts may be sent to any member of the Editorial Board. Appropriate subjects will include:

(*a*) generalized theories;
(*b*) theories of specific processes or phenomena, e.g. crossing over, learning, active transport, photosynthesis, carcinogenesis;
(*c*) theoretical discussion of specific projects, e.g. selective mutation, drug design, electronic models;
(*d*) theoretical discussion of methods, e.g. cytochemical methods and instrumentation, amplification techniques.

Papers that are concerned mainly with experimental work will not normally be accepted, though inclusion of new experimental work will be permitted where this is essential to a theoretical discussion. Reviews will be accepted only if they include a substantial new theoretical discussion. Letters to the editor, intended to direct attention to a particular theoretical point, will be accepted.

If this journal is to serve its intended purpose, i.e. to act as a *central* medium for theoretical discussion, it is essential that the major part of its contents should be readily intelligible to all biologists. To this end, wherever possible mathematical development should be confined to appendices, and all specialized terms should be clearly defined in the introduction to a paper.

J. F. Danielli

J. Theoret. Biol. (1961) 1, 1–17

Possible Functions of Chains of Catalysts

R. J. P. Williams

Wadham College, Oxford, England

(*Received* 7 *August* 1960)

A biological system utilizes chains of catalysts to connect the oxidizing power of molecular oxygen with the reducing power of organic substracts (the electron transport chain) and to connect the absorption of light with the reduction of carbon dioxide and the evolution of oxygen (the chloroplasts). The complexity of these systems is analysed in order to appreciate their requirement for an apparent superfluity of catalysts.

Introduction

At first sight a biological system appears to be unnecessarily complex. A simple example will bring out this point. In the utilization of oxygen by many biological systems there is a chain of about ten catalysts before the oxidizing equivalents of the oxygen molecule are allowed to react with the reducing equivalents of some carbon compound. This chain of electron transport separates oxidizing power from reducing power. For mitochondria from animal sources it is frequently written

$$\text{(oxidizing equivalents) } O_2 \rightarrow \text{cyt. a} \rightarrow \text{cyt. c} \rightarrow \text{cyt. } c_1 \rightarrow \text{cyt. b} \rightarrow \text{flavoprotein} \rightarrow \text{H (reducing equivalents)}$$

where cyt. stands for cytochrome. The chain of catalysts given here underrepresents the complexity of the system, for between the various steps of the chain there may be additional catalysts. Some of these are ill-defined, for example cytochrome a_3 and the copper compounds of cytochrome oxidase between O_2 and cytochrome c, and the various quinones which may link several of the components of the chain. For details of the chain see Green (1959) and Chance (1957). In the flavoproteins there are often two catalysts, a flavin and a metal ion such as molybdenum or iron, and beyond the flavoprotein there is frequently a DPN†-dependent dehydrogenase. In other living organisms a different series of cytochromes links oxygen and the same substrates so that there does not appear to be anything highly specific about the individual character of the catalysts found in the animal mitochondrion electron transport chain. The feature common to the different electron transport chains is that in order that oxygen may

† Diphosphopyridine nucleotide.

oxidize lactate or malate (two of the very many compounds it coul oxidize) there are about ten intermediate catalysts. It is surprising from naïve point of view that there are not just two. One to react with oxyge and one to react with the organic substrate, lactate or malate. Now if w believe that biological systems have become so complex in order tha special ends of a biological system may be achieved then we need t explain this long chain of catalysts. This paper is an attempt to rationaliz the advantages of such a chain.

First we list some of the features of a chain of catalysts which would b absent from a reaction between the reducing and oxidizing equivalents a a single catalytic site.

1. The catalysts—enzymes—need to interact in a specific way with thei substrates if the high selectivity required of biological systems is to b achieved. This suggests that the most effective catalysts must have tw sites, one for each of the reacting species. Thus in the biologically im portant reaction of oxygen with phenols the catalysts are metal ions an an enzyme. The reduced metal ion is thought to react with the oxygen an the protein with the phenol. This type of system fails to explain the lengt of the electron transport chain, though it introduces specificity in regar to the phenol oxidized.

2. It may be necessary for several reagents A to react with a give substrate B, while specificity is required for all of the reagents A. Fo example, very many oxidations are carried out by molecular oxygen, her B. The oxygen is absorbed on a given site, cytochrome oxidase. By con necting this site with a number of other catalytic sites rather than wit just one, a variety of different reagents can be oxidized without a multi plicity of sites for oxygen absorption. This is the familiar branched electron transport chain.

$$O_2 \rightarrow \text{cytochromes} \begin{matrix} \nearrow \text{succinate} \\ \rightarrow \text{malate} \end{matrix}$$

It clearly requires at least cytochrome oxidase, a series of dehydrogenases, and some freely diffusing entities which carry the oxidizing and reducing equivalents between the two groups of catalysts if these are not in contact. This explanation of the value of the chain does not explain why it is necessary to have several non-diffusible intermediate catalysts before there is any branching in the chain, e.g. cytochromes a_3 (?), a, c, c_1, plus possibly copper and some quinones.

3. A series of catalysts can be used to package energy. Thus in the cytochrome chain there are a number of steps of somewhat similar redox potential difference. It could be that there is a convenient energy packet which the biological system can readily handle. This explanation of a

series of catalysts does not indicate why both cytochromes c and c_1 are present in the chain, for their separation in redox potential is trivial. Nor will it explain why cytochrome a_3 (?) cytochrome a and copper are required for cytochrome oxidase.

4. A series of catalysts serves to separate in space the substrates A and B. Obviously the reactive site for oxygen, cytochrome oxidase, is some distance from the reactive sites, the dehydrogenases, for the organic substrates, malate and succinate. Activated oxygen does not come physically near (in space) to activated organic substrates. This does not imply that oxygen and these substrates are diffusion-restricted. Only the sites of activation are rigidly located in space. However, because the sites of reaction are separated there will be a time lag, the time of diffusion, before the products at the two reaction sites can come together. There are two important controls upon reaction then: activation of substrates, and diffusion of products. It may well be that the achievement of a separation of activated reagents in space plus restricted diffusion provides the fundamental distinction between biological chemistry and test-tube chemistry. In the reactions carried out in a test-tube without catalysts diffusion of all reacting species is permitted; although the course of a reaction can be controlled by the addition of a suitable catalyst which activates a molecule in a particular way, e.g. the dehydration of alcohol can be made to go in one of two directions

$$CH_3CH_2OH \begin{cases} \nearrow CH_2CH_2 + H_2O \\ \searrow CH_3CH_2\,OCH_2CH_3 + H_2O \end{cases}$$

it is not possible to control reactions in a test-tube at room temperature if the reactions are of low activation-free energy, i.e. reactions that do not require catalysts. For example, the reaction $H^+ + OH^- \rightarrow H_2O$ is virtually instantaneous in free solution and is equally probable in all parts of space. All parts of such a system are held at the same pH, the same hydrogen ion activity. The situation differs entirely from that in the case where diffusion is restricted as in an electrolysis experiment using a cell without transference. The two electrode compartments are separated by a salt bridge. Consider the cell

			salt bridge		
Pt	Fe^3X^- Fe^2			H^+ H_2	Pt
	pH 5·0				

in which all ions are 0·1 molar and where X is any ligand for which H^+ has a high affinity, e.g. ethylenediamine tetra-acetate. The overall reaction $\frac{1}{2}H_2 + Fe^3X^- \rightarrow H^+ + Fe^2 + X^-$ takes place but not the further reaction $H^+ + X^- \rightarrow HX$ which would occur in the absence of the salt bridge. The effect of the salt bridge then is to prevent the coming together of

some of the species (H^+ and X^-) which would otherwise react. This type of cell is particularly significant as one compartment contains the hydrogen ion at an entirely different activity from the other. In some biological systems a similar situation is realized. Activity of a substance is only meaningful in a particular spatial region of the biological system. However, the restriction of a reaction by preventing mixing is only one way of controlling reactions in space. For substances which need catalysts in order to react, restriction of diffusion by a salt bridge or a membrane is unnecessary and their "reactivity" can be controlled in space by the manner of assembly of a set of catalysts. Where reactive entities C and D are produced in separate parts of space by such catalysts, immediate subsequent catalysed or uncatalysed reaction with other groups, Y and X respectively, can be used to prevent their reaction with one another. They can be made to react by suitably placed catalysts before they can diffuse to one another although diffusion itself need not be restricted. The scheme is

generation from B of a product, C, at a particular site followed by $Y + C \xrightarrow{\text{catalyst}} CY$	catalytic chain	generation at a different site of D by reaction of A followed by $D + X \xrightarrow{\text{catalyst}} DX$

The reaction $C + D \rightarrow CD$ is not permitted (or very restricted) because reactants C and D cannot come together quickly enough or because the reaction $CY + DX \rightarrow X + Y + CD$ has no catalyst. It is quite possible that the reaction of oxygen with organic substrates in a biological system is of this kind. Oxygen and the small organic molecules that are oxidized by it, e.g. succinate and malate, can diffuse rapidly in biological systems, and their reactions are controlled by catalysts far apart in space. One might then ask what are the possible substances completing the scheme

generation from oxygen of a product C, followed by $C + Y \rightarrow CY$	electron transport chain	generation from reduced organic matter (malate, succinate) of D followed by $D + X \rightarrow DX$

Here we will later identify C with OH^-, Y with phosphorylation of sugars, D with H^+, and X with ADP + P†. Before we can reach this stage of speculation we must first ask further questions about the chains of catalysts and about the processes which a biological system carries out.

One of the points we have not examined yet is the reason for the complexity of the co-enzymes which form part of the electron transport chain. Biological catalysts, enzymes, are large molecules which cannot diffuse

† Adenosine diphosphate and Phosphate.

rapidly over great distances through a system that is partly aqueous and partly non-aqueous. The co-enzymes too are so large and so tightly bound by their enzymes that they must be almost permanently restricted to special parts of space. It is known for example that even some cations, which are the smallest co-enzymes (cofactors), e.g. zinc in carboxypeptidase, do not exchange rapidly with the surrounding aqueous medium (Vallee, 1955). However, some groups must be used to transfer energy, electrons or hydrogen, from one member of the chain to another. Possibly these are metal ions, e.g. Fe^{II} ions and the proton, acting merely as redox

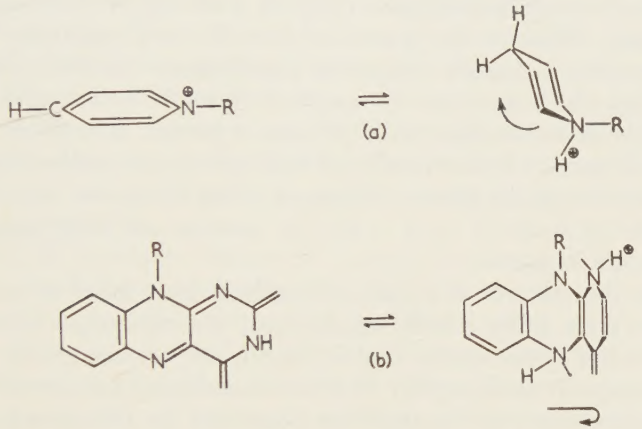


FIG. 1. The swing of a co-enzyme, depicted by an arrow, and illustrated for DPN (*a*) and a flavin (*b*). In each case the group R–N acts as the fixed part of the co-enzyme about which the rest of the molecule swings (Velick, 1958). Group-transferring co-enzymes could well work by similar mechanisms so that there is no need for the migration of the enzyme or co-enzyme as a whole.

mediators, which diffuse between certain catalysts, proteins, that would react but cannot come into physical contact with one another. Now while it is easy to appreciate that a combination of a diffusable co-enzyme or co-factor and a protein have complementary functions and therefore there is a need for both, it is not so easy to appreciate the requirement for co-enzymes which are large molecules. In such a case neither protein nor co-enzyme will diffuse rapidly. Let us examine what purpose could be fulfilled by the complex structure of the co-enzymes, e.g. of DPN. The pyridine ring moiety of DPN carries out the redox reactions but it seems probable that the pyro-phosphate group only attaches the co-enzyme to the enzyme: for example it may well bind DPN to zinc of alcohol dehydrogenases. During oxidation and reduction cycles these co-enzyme molecules go from a benzenoid to a quininoid ring structure and back again (Fig. 1).

This type of reaction will be accompanied by considerable steric changes in the co-enzyme. Its movements can be likened to the opening and closing of a door, the molecule hinging about the bisector of the heterocyclic ring which passes through the pyridine nitrogen atom. The phosphate group serves as the hinge about which the whole co-enzyme swings. Such a swinging movement has a clear significance in that it will lead to the transport of the reactive group in a specific direction—perhaps from one protein to another, increasing the distance between the site of initial attack and that of the final production of products. It is most noticeable that many co-enzymes have the ability to undergo this type of swing, for they are usually made of two distinct parts (Fig. 1). The way in which they swing varies greatly. Thus we can appreciate that the very complexity of a co-enzyme provides a definite directional significance—the direction of permitted swing which could not be supplied efficiently by a simple diffusing ion. A series of such swings could produce a pattern of reaction in space. The over-all pattern is the transfer of hydrogen by the dehydrogenases to the last member of the electron transport chain where we have, at a particular point in space $\oplus + H \rightarrow H^+$, i.e. protons are being generated at specific points in space.

Whereas the features of a chain of catalysts introduced by points 1, 2 and 3 above are fairly readily appreciated, the advantages which could accrue to a biological system under point 4 are not necessarily obvious. For this reason we shall amplify what we have already said about reactions in which products and reactants are restricted by the structure of the catalyst chain to particular different regions of space—we shall call them *dislocated* reactions. In order to appreciate the importance of such reactions we must first look at the type of reaction occurring in a biological system.

The Reactions of Biological Systems

The thermodynamically stable form of carbon/hydrogen/oxygen mixtures is carbon dioxide and water. The fact that the elements are observed in much more complicated compounds, especially in biological systems, implies that energy has been put into the intercombinations. The great majority of these compounds are kinetically, although not thermodynamically, stable either in the presence or absence of oxygen at room temperature. The energy that has been put into the systems and continues to be put into them is radiant energy from the sun. It has been shown recently that an early step in the conversion of light energy into chemical energy is the formation of ATP† (Losada, Trebet, Ogata & Arnon, 1960). Later ATP reacts to give ADP and the energy gained can be used in assimilating

† Adenosine triphosphate.

carbon dioxide into reduced compounds such as glycollic acid. Glycollic acid is one of the main (unstable) two-carbon fragments from which the more complicated (but still unstable) compounds of biology are formed. This then is a major synthetic path involving reduction of carbon dioxide. A second synthetic path utilizes acetate. Acetate, a two-carbon fragment, can be thought of as containing one largely reduced carbon, CH_3— and one largely oxidized carbon, —CO_2^-. The oxidizing and reducing equivalents in acetate can be rearranged by catalysts in a number of ways so that the degrees of reduction or oxidation (oxidation states) of the carbon atoms in the final molecule are more equal, e.g. 3 $CH_3CO_2H \rightarrow C_6H_{12}O_6$. This process requires energy, since glucose is unstable with respect to acetate, and ATP is again the source of this energy. Some of the intermediate compounds formed are citrate, malate, and citramalate (Losada *et al.*, 1960). These molecules can be used in further synthesis of the large unstable molecules. A second way in which the essential ATP can be formed is by oxidative phosphorylation. Thus we have two synthetic paths (there may be many others), one aerobic, using oxygen, and the other anaerobic, using light.

The second important feature of biological systems is that they also carry out controlled degradation of carbon compounds such as sugars. The controlled oxidation may occur aerobically, using oxygen when it normally yields carbonic acid as a final product, or anaerobically when a disproportionation of carbon compounds occurs so as to produce simultaneously more reduced and more oxidized carbon atoms either in two separate compounds, e.g. in fermentation of sugars to alcohol and CO_2, or in the same compound, as in fermentation to acetic acid. This last reaction represents a redistribution of oxidizing and reducing equivalents amongst the carbon atoms in one molecule, i.e. in acetate, so as to produce wide disparity between them. The two degradation reactions are the reverse of the two synthetic paths. They are coupled with the re-synthesis of ATP from ADP, as the reactions are now thermodynamically favourable.

Before proceeding with a discussion of the character of these syntheses and degradations we should note the quantitative energy relationships between the component molecules of the paths. Figure 2 is provided in order to illustrate the way in which the thermodynamic stability of a C/H/O molecule is related to the distribution of oxidizing equivalents amongst the different carbon atoms. The number of oxidizing equivalents in the oxidation state of a molecule, is calculated using the following system (cf. Rabinowitch, 1951). Each atom of oxygen in a compound $\equiv +2$, each atom of hydrogen $\equiv -1$, each negative charge $\equiv -1$, each positive charge $\equiv +1$. Thus the value for CO_2 is $+4$, for CH_4 is -4 and for $C_6H_{12}O_6$, $CH_3CO_3^-$ or CH_3COOH is zero. The oxidation state

of a molecule is plotted on the abscissae in the figure. The ordinate is the standard free energy of formation of the aqueous solution (pH 7·0) of the compound from the elements and is given per carbon atom in the compound. The value of ΔG in the figure for glucose is therefore 219·2/6 or 36.5 kcal per carbon. Values for the standard free energy per mole of aqueous solutions, e.g. for glucose, 219·2 kcal, can be obtained from

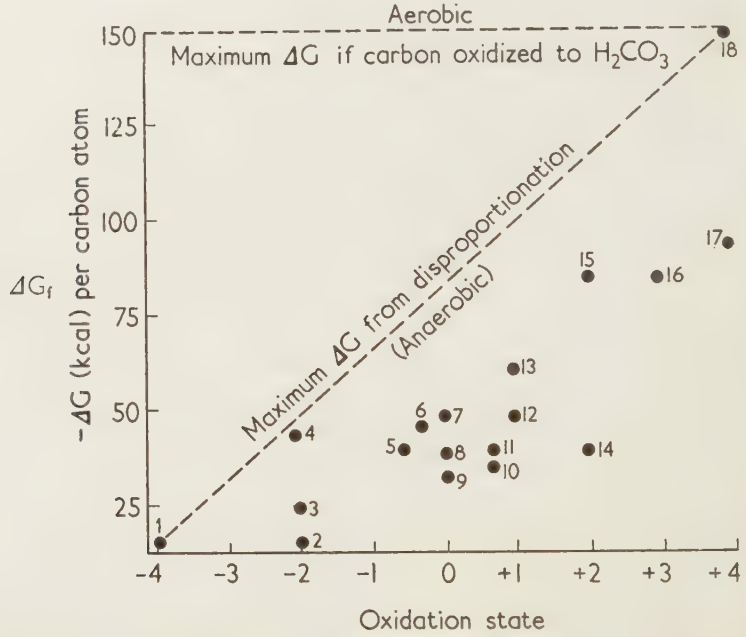


Fig. 2. Standard free energy of formation from the elements per carbon atom of a series of carbon compounds plotted against oxidation number. The numbers refer to: (1) methane; (2) *n*-butanol; (3) ethanol; (4) methanol; (5) glycerol; (6) citrate; (7) acetic acid; (8) glucose; (9) formic acid and aceto-acetate; (10) pyruvate; (11) keto-glutarate; (12) malate; (13) glycollate; (14) carbon monoxide; (15) formic acid; (16) oxalate; (17) carbon dioxide; (18) carbonic acid.

"Energy Transformations in Living Matter", by H. A. Krebs and H. L. Kornberg, Springer-Verlag, Berlin, 1957. The following points are clear from the graph.

(i) Unsaturated systems are less stable than saturated ones.

(ii) Acetate is more stable than glucose although both have the same oxidation number.

(iii) Disproportionation of any carbon compound to H_2CO_3 (aq) and methane is always thermodynamically favourable as is disproportionation to alcohol (methyl or ethyl) and carbonic acid. The

pattern of reaction here is anaerobic—no new oxidizing equivalents are added to or removed from the system.

(iv) Under aerobic conditions all carbon compounds tend to go to carbon acid, H_2CO_3, i.e. CO_2 and H_2O, unless free energy can be retained by the system for synthetic purposes.

To summarize these observations: all biological systems have a natural tendency to degrade, redistributing the oxidizing equivalent so as to produce carbon in -4 or $+4$ oxidizing states. It is a sufficient condition for degradation reactions that catalysts or energy are supplied for activation. However, in order to achieve synthesis it is necessary to supply not only energy but also catalysts that allow the synthesis and prevent the degradation; synthesis is not just a reversal of degradation.

Now it is known that the redistribution of oxidizing and reducing equivalents amongst components of biological systems occurs only with the small molecules. Once a large molecule is built by condensation between the small units, oxidation-reduction reactions occur infrequently and the general pattern of reaction is continued condensation—e.g. polysaccharides, proteins, nucleic acids.

A further feature of a biological system is that it carries out simultaneously the two sets of reactions, natural degradation and biochemical synthesis, by redistributing oxidizing and reducing equivalents amongst carbon atoms. We must note especially that the synthesis and the degradation occur together in the same vessel and particularly "wanted" compounds are being synthesized at the expense of the destruction of molecules that are unwanted but differ inappreciably on a thermodynamic scale of stability from the wanted compounds.

In less technical language it is a feature of many biological systems that certain "food-stuffs" which are part of a living organism are degraded in order to make the parts of a second living organism. One form of life eats another form in order to grow. This type of process will not occur if the living organism fails to distinguish between itself and its food. There must be an arrangement by which it is possible to say that this molecule will not be degraded but will be incorporated into the synthetic path. At the level of big molecules those which we call foodstuffs are sufficiently different from those we consider to be part of the living organism to be readily distinguished by other big molecules. However, at the level of small molecules, this distinction disappears, as the units from which the complicated proteins and sugars, etc., are built are the same whether the proteins or sugars are food or part of the living organism.

We recognize that a molecule such as a simple triose or a simple amino acid cannot be said to be a foodstuff or part of a living organism: it could

react to be either. Moreover, these molecules are small and presumably can diffuse freely to sites where synthesis or degradation is taking place. The small molecules form an *essential common pool* to the degradative and synthetic paths, for these paths cannot be separated either chemically or physically at the lowest level of chemical complexity. The units which are used in synthesis and degradation are not separated physically in space (i.e. diffusion is not restricted) but rather they are trapped on certain active sites which are common to both paths. Only when these units are built into larger molecules can they be said to be incorporated in the synthetic path. We shall now show that it is a feature of the known reactions of the common pool that they are neither wholly degradative nor wholly synthetic but are both, so that small molecules can be switched from one path to the other.

Any set of reversible reactions is both synthetic and degradative but it will only synthesize the same set of substances as it degrades. On the other hand a cyclic set of redox reactions is much more flexible for, in order to return to the same redox state as that from which it starts, it must contain both degradative and synthetic steps. The differences between reversible series of reactions and cyclic series are illustrated in Tables 1 and 2. The Kreb's cycle contains both synthetic and degradative steps whereas the degradation of sugars is a series of degradative steps only. Although a series of reversible reactions can be made to go in either direction by the controlled supply of initial reactants it will not degrade and synthesize at the same time. Perhaps this difference explains why cyclic series of

TABLE 1

Oxidation states of the Kreb's cycle

Reaction substrates	Oxidizing equiv. added	No. of carbon atoms of oxidation state								
		4	−3	−2	−1	0	+1	+2	+3	+4
Acetate + oxalo-acetate	0		1	1				1	3	
Citrate	0			2			1		3	
cis-Aconitate	0			1	1	1			3	
iso-Citrate	0			2			1		3	
α-Keto-glutarate + CO_2	2			2				1	2	1
Succinate + $2CO_2$	2			2					2	2
Fumarate + $2CO_2$	2				2				2	2
Malate + $2CO_2$	0			1		1			2	2
Oxalo-acetate + $2CO_2$	2			1				1	2	2

The effect of the first two steps is to make the distribution of oxidizing equivalents more even over the carbon atoms; the steps are synthetic. The later steps increase the disparity in the distribution of oxidizing equivalents or increase the total number, i.e. they are degradative steps.

TABLE 2

Oxidation states in sugar degradation

Reaction substrates	Oxidizing equiv. added	No. of carbon atoms of oxidation state								
		−4	−3	−2	−1	0	+1	+2	+3	+4
Glucose	0				1	4	1			
Glyceraldehyde	0				2	2	2			
Dihydroxyacetone	0				4			2		
Glyceric acid	4				2	2			2	
Enol-pyruvic acid	0			2			2		2	
Pyruvic acid	0		2					2	2	
Lactic acid	0		2			2			2	

All steps to pyruvic acid are degradative. Lactic acid is added to illustrate a synthetic step.

reactions are so important in biology. However, while the cycle of reactions turns over, energy must be conserved so that the further steps of synthesis outside the cyclic reactions of the small molecules can be achieved. Moreover this energy must be conserved and transferred without altering disparately the oxidation states of the carbon atoms in the synthetic path. We will now examine this energy transfer more closely to see how it can occur.

ENERGY TRANSFER

The problem is how to transfer energy in these systems in order to get the simultaneous synthesis of some steps and degradation of the others to occur in a coupled manner. The reactions in the cycle must be coupled, as the synthetic steps are energetically unfavourable; the free energy evolved in the degradation steps must not be lost as heat but must be retained for the syntheses of the other steps of the cycle. Also the reactions are monitored by phosphorylation reactions: excess free energy is trapped in pyrophosphate bonds, ATP. The significance of the phosphorylations is understood better when their occurrence is more generally examined.

As we have stressed in this article, a biological system requires energy. Energy is put in primarily through light absorption. As light absorption has led historically to the production of free oxygen there is at the present time a second path for energy absorption which depends ultimately upon light absorption: this is the absorption of oxygen. We shall now consider the part played by phosphorylation in both these paths,

$$\text{light and system A} \rightarrow \text{system B} + \text{growth} \; (+ \; O_2), \text{ and}$$
$$\text{oxygen} + \text{system B} \rightarrow \text{growth} + CO_2.$$

The absorption of oxygen

This path introduces new oxidizing equivalents per carbon into system B. Yet we have seen that the synthetic path required for growth does not involve the addition of new oxidizing equivalents, but builds up sugar and proteins from acetate without changing the number of these equivalents per carbon. These oxidizing equivalents are expended in the degradation reactions which yield CO_2. Again we know that the energy supplied by oxygen is not transferred to the synthetic path through oxidation-reactions but through oxidative phosphorylation. In fact energy must not be transferred by oxidation-reduction reactions between the paths or degradation, increase in the number of oxidizing equivalents per carbon atom, would be bound to occur and there could be no synthesis. The medium for the transfer of the energy of the reaction between carbon compounds and oxygen must not be a redox catalyst for it transfers oxidizing equivalents from substrate to substrate, only increasing or decreasing the degree of oxidation or reduction in accord with the nature of the added substrate, i.e. depending on whether it is oxidizing or reducing. In aerobic reactions oxidation must be kept separate from synthesis so that oxidizing equivalents cannot enter the synthetic path.

The following conditions must be satisfied:

(*a*) The known initial attack of oxygen is at cytochrome oxidase which is believed to result in the generation of electron transport in the cytochrome electron transport chain (p. 1).

(*b*) The electron which is transported in the chain comes from the reaction $H + \oplus \rightarrow H^+ + e$. This reaction occurs where the electron transport chain meets the hydrogen atom transport chain of the dehydrogenases (p. 6).

(*c*) The original oxygen molecule must pick up negative charge, and, reacting with water, give hydroxyl ions.

(*d*) The energy of the over-all oxidation reaction of oxygen with the hydrogen from organic substrates is not lost as heat but is used to generate a condensation of phosphate in the reaction $ADP + P + H^+ \rightarrow ATP$. Thus the proton generated in stage (*b*) is consumed in a condensation reaction, stage (*d*).

The combination of these steps of reaction succeeds in switching energy of oxidation reactions into energy of condensation reactions. We have stressed that such a switch was essential for the coupling of synthesis and degradation. The scheme for aerobic reactions is then

$$O_2 \rightarrow \text{cytochrome chain} \begin{cases} \uparrow \rightarrow H\oplus \leftarrow \text{reduced dehydrogenases.} \\ \downarrow \\ OH^- \end{cases}$$

The generated protons and hydroxyl ions are separated in space, and the oxidizing power of oxygen has been converted into the energy of ionization of the water molecule. Now changes in concentration of the proton can be used to bring about the reaction

$$\mathrm{ADP} + \mathrm{P} \xrightarrow{\mathrm{H^+}} \mathrm{ATP}$$

for this reaction is thermodynamically favourable at *low* pH. If, as we suggest, the proton is produced in high *local* concentration in the region of the ATPase then this absorption of the proton in the polymerization of phosphate will prevent the energetically more favourable $H^+ + OH^- \rightarrow H_2O$. Moreover, this reaction will not occur subsequently if the OH^- ion is already absorbed in another manner at another place. We can suggest that this reaction is:

$$\text{sugar} + \mathrm{ATP} \rightarrow \mathrm{ADP} + \text{sugarphosphate} + \mathrm{H^+}$$

which will be driven forward if the OH^- ion is added to remove H^+. Thus the OH^- ion phosphorylates sugars while the H^+ ion can bring about oxidative phosphorylation in another part of space. The phosphorylated sugar then enters into the set of degradative reactions which further stimulates the generation of ATP.

	O_2 ↓	
phosphorylation ← OH^- of sugars	electron transport chain	H^+ → oxidative phosphorylation

Such a scheme of *dislocated* reactions can only occur if there is the correct assembly of catalysts, and is therefore a feature of a highly organized system, such as a biological system. The scheme of p. 4 is now finalized for oxidative phosphorylation.

The oxygen atom of OH^- here comes from the oxygen molecule and the hydrogen of H^+ from the dehydrogenases or eventually from carbon compounds. These groups or ions are not present in equilibrium concentration if the assembly is considered as a whole.

Energy from light

The action of light, if not controlled, can only be the addition of activation energy to an unstable system. The system will tend to go then to the most thermodynamically stable products unless the energy is incorporated in some subtle manner. Light absorbed amongst carbon compounds should act only to disproportionate acetate to carbonic acid and alcohol or methane. This is the anaerobic degradation reaction. But if the energy

from the light is to be incorporated into the synthetic path we must avoid such a reshuffle of oxidizing equivalents. It must be prevented for the same reason as the increase in oxidizing equivalents had to be prevented in the aerobic path. The large molecules of biological significance have an average oxidative number of roughly zero and its distribution over the carbon atoms is rather even. The over-all light reaction without CO_2 uptake is $CH_3CO_2H \rightarrow 2\ (CH_2O)$.

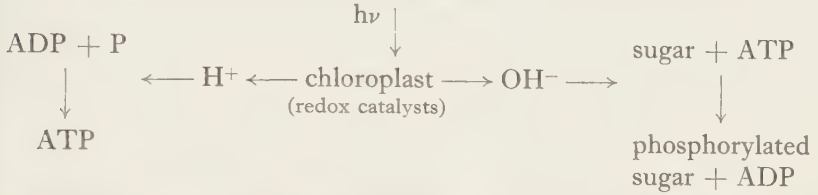
The light is to be used to reverse disproportionation amongst oxidizing equivalents, not to cause it: the transfer of energy from the degradative path for synthetic purposes through redox catalysts must not be allowed, for such transfer could only result in a change of oxidation state according to the degradative path, and cannot produce synthesis.

Now the second feature of photosynthesis which is similar to oxidative phosphorylation is that it involves a great number of catalysts. The photosynthetic unit (Hill & Wittingham, 1956; Rabinowitch, 1951) involves a complexity of pigment molecules of the chlorophyll and carotenoid kind and a variety of cytochromes. Much as is the case with mitochondria, there is believed to be a regular array of these units but there is nothing highly specific about the individual members. The photosynthetic bacteria use slightly different photo-active compounds from those of the photo-active plants. There is also a variety of cytochromes involved. The common feature to all the photosynthetic units is the number of the catalysts involved which appears unnecessarily large for the reactions that are carried out, and it has been repeatedly postulated that transport of the energy of the absorbed light takes place over a considerable distance in space. The complexity of the system makes it possible for the immediate interaction of the light with the active pigment to result in a chemical change some distance away from the initial site of attack.

The work of Chance & Olson (1960) shows that a very early, physically observable step in the chain of reactions subsequent to the absorption of light is the oxidation of a cytochrome. Thus the light energy absorbed in chlorophyll is transferred early in the process of photosynthesis to a second catalytic system and is not simply utilized for chemical reaction by the catalyst absorbing it. The photo-synthetic unit is a series of catalysts essentially similar to the unit of the electron transport in mitochondria.

The work of Losada *et al.* (1960) strongly indicates that another early step in photosynthesis is the formation of ATP from ADP. Thus we have a suggestive similarity between mitochondria and chloroplasts. The energy that is required by the system is rapidly transferred to a cytochrome system some distance from the initial site of attack, either by oxygen or light, and this energy is then transferred, probably immediately, to ATP,

the last step being essential if the system is to carry out synthesis. We can indicate the photosynthetic scheme diagrammatically.



Now the subsequent reactions of this system, which is identical to that given for the mitochondria, depend upon the catalysts available for utilization of sugar phosphates. In mitochondria the sugar phosphates enter the degradative path, for only in this way can the production of ATP be achieved. The sugar is regarded as the ultimate source of energy in its oxidation by molecular oxygen. It supplies the hydrogen shown as $H \rightarrow H^-$ in the scheme on p. 13. This set of reactions is absent in a photosynthetic unit that can reduce compounds while generating oxygen. It can therefore have no catalysts to use molecular oxygen in oxidation reactions in this region of space; for, otherwise, degradation, which is thermodynamically favourable, would inevitably occur as the overall process. The sugar phosphates are used instead in the production of starch. Where ATP is made available the energy necessary for this further synthesis of sugar phosphate to starch is present in the system. This reaction will be able to continue so long as there is a supply of light and a supply of carbon. The nature of the supply of carbon is not material once ATP is supplied; it can be acetate or carbon dioxide, as either of these materials can be supplied to reverse degradation reactions. It is well established that many degradation reactions are reversible (Vishniac & Ochoa, 1951; Losada *et al.*, 1960) e.g.

$$\begin{array}{l} CH_3 \\ | \\ CO \\ | \\ CO_2H \end{array} + CO_2 \rightarrow \text{oxalacetic acid, the Wood \& Werkman reaction,}$$

and

$$CO_2 + \text{pyruvic acid} \rightarrow \text{malic acid}$$

and

$$CO_2 + \alpha\text{-ketoglutamic acid} \rightarrow \text{isocitric acid.}$$

More nearly similar to the photochemical reaction sequence is the set of reactions in which assimilation occurs with phosphorylation as in the reactions:

Pyruvate + ATP → phospho-enol-pyruvate + ADP
Phospho-enol-pyruvate + CO_2 → oxalo-acetate phosphate.

(This set of reactions is the set which normally runs in the reverse direction in the Kreb's cycle so as to conserve energy and not to assimilate.)

In this reaction the important point is illustrated that the uptake of CO_2 will require large quantities of energy. This energy can be obtained from the reaction ATP → ADP + P.

The photochemical sequence is:

Ribose phosphate + ATP → ribulose-diphosphate + ADP
Ribulose diphosphate + CO_2 → two mols phospho-glyceric acid
Phospho-glyceric acid → glucose + phosphate

Thus the production of polysaccharides is associated with the reaction

$$ATP \rightarrow ADP + P + H^+$$

which generates protons at pH 7·0.

The series of reactions is seen to be the generation of ATP by the production of H^+ in one locality and its degradation in a second locality by the reaction of OH^-.

The reversal of degradation reactions further leads to the incorporation of hydrogen into carbon compounds. Chance & Olson (1960) have shown that this is a subsequent step of photosynthesis. The hydrogen frequently originates from water in photosynthesis although other hydrogen donors such as H_2S can be used. These hydrogen donating reactions are responsible for the production of oxygen, photosynthesis, or sulphur, in the sulphur-generating photochemically active bacteria. The source of the hydrogen is incidental to the photosynthetic process and the generation of oxygen is also only a result of the use of water as this source of hydrogen. It could well be that the initial reactions of chloroplasts are very similar to the dislocated reactions postulated in mitochondria.

Conclusion

In this account we have discussed the catalytic chains of biological systems, indicating the importance of the properties which, we consider, such a chain can have. We have elaborated a special feature of such a chain—that it allows the use of space in dislocating chemical reactions. That such a utilization is common to many biological systems is obvious enough when we think of such familiar phenomena as sense perception. Perhaps the complexity of the latter systems has developed from a fundamental feature at the lowest level of biological synthesis—the utilization of organization in chemistry. While all such arguments are speculative there is a problem to which the arguments are addressed. Why does a

biological system need so many energy- and electron-transporting catalysts? Our answer—so that the system may make use of space—removes the necessity of postulating semi-permeable membranes in many biological systems. A catalytic chain of the kind described here could have all the functional significance normally ascribed to a semi-permeable membrane.

REFERENCES

CHANCE, B. (1957). *Advances in Enzymology*, **17**, 65.
CHANCE, B. & OLSON, J. M. (1960). *Arch. Biochem. Biophys.* **88**, 54.
GREEN, D. E. (1959). *Advances in Enzymology*, **21**, 73.
HILL, R. & WITTINGHAM, C. P. (1956). "Photosynthesis". Methuen, London.
LOSADA, M., TREBET, A. V., OGATA, S. & ARNON, D. I. (1960). *Nature, Lond.* **186**, 753.
RABINOWITCH, E. I. (1951). "Photosynthesis", 2nd Edition. Interscience, New York.
VALLEE, B. L. (1955). "Advances in Protein Chemistry", **10**, 317. Academic Press, New York.
VELICK, S. F. (1958). *J. biol. Chem.* **233**, 1455.
VISHNIAC, W. & OCHOA, S. (1951). *Nature, Lond.* **167**, 768.

J. Theoret. Biol. (1961) **1**, 18–26

Permeation through a Spherical Membrane

G. G. Hall

Mathematics Department, Imperial College, London, England

(*Received* 23 *August* 1960)

The diffusion of matter from a cell is considered in terms of a model in which a fixed volume is bounded by a thin permeable membrane. The equations governing the permeation of matter through this membrane and its subsequent diffusion are derived. Their solution involves several new Laplace transforms. The variation of concentration inside and just outside the cell is treated in detail. Graphs of these concentrations are given for the particular example of haemolysing erythrocytes.

1. Introduction

This paper considers a simple model of a cell consisting of a thin permeable membrane enclosing a sphere of fixed volume. This simple model is used to investigate some of the problems of diffusion from a cell. It is too simple to reproduce phenomena which depend on the internal features of the cell or the detailed structure of its bounding membrane, but it should be sufficient to enable some of the stages of the diffusion process to be distinguished and some of the consequent effects to be estimated. The results of greatest importance are not very sensitive to the spherical shape assumed in the model in order to facilitate calculation. In the example which is treated in more detail, that of haemolysis from red cells *in vitro*, the cell in fact does swell to a sphere before the membrane becomes permeable to the haemoglobin.

It is assumed that diffusion inside the cell is sufficiently rapid to maintain a uniform concentration of material there. The permeable membrane acts as a barrier and so determines the rate of the whole process. Once the material has penetrated outside it begins to diffuse fairly rapidly into the surrounding medium until the concentration is uniform. The macroscopic equations are here solved so that the concentration can be found at any point and any time. Of especial importance is the material which has escaped from the cell but not had time to diffuse away. This material will give rise to a back-effect which slows down the permeation and it may also interfere with observation of the cell.

2. The Mathematical Model

The essential features of the model, mathematically, are that a sphere, of radius a, contains a uniform concentration $c(t)$ of material and that its surface is a permeable membrane of thickness d. The permeability constant P is defined by the equation

$$\text{rate of matter out} = PA\,\Delta c/d,$$

where A is the surface area and Δc the difference in concentrations inside and out. The material outside diffuses away with diffusion constant D. The concentration outside is a function of the radial distance from the cell centre and the time and satisfies the usual diffusion equation

$$\frac{\partial^2 f(r,t)}{\partial r^2} + \frac{2}{r}\frac{\partial f(r,t)}{\partial r} = \frac{1}{D}\frac{\partial f(r,t)}{\partial t} \tag{2.1}$$

Initially, at $t = 0$, the matter is entirely inside the sphere so that

$$c(0) = c_0;\ f(r,0) = 0, \text{ all } r. \tag{2.2}$$

Finally, the material is uniformly dispersed throughout the volume available and, since this is unlimited, the concentration vanishes everywhere, i.e.

$$c(\infty) = 0;\ f(r,\infty) = 0, \text{ all } r. \tag{2.3}$$

In these terms, the problem is to find the equation determining $c(t)$ and then to solve the diffusion equation subject to the initial conditions and to the boundary conditions implied by the permeation equation.

There are close analogies between this problem and that of diffusion through three regions having different diffusion constants. This problem has been studied by various authors (Auer & Murbach, 1954; Berthier, 1955).

3. Permeation Through the Surface

As the material permeates the spherical shell its concentration inside should, in principle, depend on position but, in practice, if the sphere is small and the surface not very permeable it will be a very good approximation to take the concentration inside as depending on time alone. With this simplification the solution of the permeation problem is not difficult.

The definition of the permeability constant P leads directly to an equation for $\frac{dc}{dt}$ viz.

$$-\frac{4\pi}{3}a^3\frac{dc}{dt} = P\frac{4\pi a^2}{d}[c - f(a,t)] \tag{3.1}$$

This equation is linear in c and can be solved by the usual methods to give

$$c = e^{-\frac{3Pt}{ad}}\left\{c_0 + \frac{3P}{ad}\int_0^t e^{\frac{3P\tau}{ad}} f(a, \tau)d\tau\right\} \tag{3.2}$$

once $f(a, t)$ is known, this equation defines c completely. The first term in the equation is the solution that would be obtained for $f(a, t) = 0$ so that the second term represents the way in which the variation of concentration outside slows down the permeation.

The permeation may also be considered in terms of the concentration outside the sphere. It is convenient to introduce the variable u, defined as

$$u(r, t) = r f(r, t) \tag{3.3}$$

to describe the concentration since it satisfies the simpler equation

$$D\frac{\partial^2 u}{\partial r^2} = \frac{\partial u}{\partial t} \tag{3.4}$$

The total amount of material outside, at any time, is given by

$$4\pi\int_a^\infty f(r, t)r^2 dr = 4\pi\int_a^\infty u(r, t)r dr \tag{3.5}$$

and, consequently, the rate of permeation must also be

$$\begin{aligned} 4\pi\frac{\partial}{\partial t}\int_a^\infty u(r, t)r dr &= 4\pi\int_a^\infty \frac{\partial u}{\partial t} r dr \\ &= 4\pi D\int_a^\infty \frac{\partial^2 u}{\partial r^2} r dr \\ &= 4\pi D\left[u(a, t) - a\frac{\partial u(a, t)}{\partial r}\right] \end{aligned} \tag{3.6}$$

This expression can be equated to (3.1) to give another relation between the inside and outside concentrations. Thus

$$P\frac{4\pi a^2}{d}\left[c - \frac{u}{a}\right] = 4\pi D\left(u - a\frac{\partial u}{\partial r}\right) \tag{3.7}$$

and, in conjunction with (3.2), this gives

$$\frac{\partial u}{\partial r} = \left(\frac{P}{dD} + \frac{1}{a}\right)u - \frac{aP}{dD}e^{-3Pt/ad}\left\{c_0 + \frac{3P}{ad}\int_0^t e^{3P\tau/ad}\frac{u}{a}d\tau\right\} \tag{3.8}$$

where both u and $\frac{\partial u}{\partial r}$ are taken at $r = a$. This equation has eliminated c and so gives the boundary condition which must be satisfied by u at $r = a$.

4. Diffusion Outside the Sphere

The problem is now reduced to the solution of the diffusion equation

$$D\frac{\partial^2 u}{\partial r^2} = \frac{\partial u}{\partial t} \tag{4.1}$$

with the boundary condition (3.8). This equation is most easily solved by means of the Laplace transform defined by

$$\bar{u}(r, p) = \int_0^\infty e^{-pt} u(r, t)dt. \tag{4.2}$$

The diffusion equation is then transformed into

$$D\frac{d^2\bar{u}}{dr^2} = p\bar{u} \tag{4.3}$$

and the solution which vanishes as $r \to \infty$ is

$$\bar{u} = A\, e^{-q(r-a)} \tag{4.4}$$

where A is a constant and

$$q = \sqrt{\frac{p}{D}} \tag{4.5}$$

The constant A is fixed by the boundary condition at $r = a$. The transform of the boundary condition (3.8) is the equation

$$\frac{d\bar{u}}{dr} = \left(\frac{P}{dD} + \frac{1}{a}\right)\bar{u} - \frac{aP}{dD}\left(c_0 + \frac{3P}{ad}\frac{\bar{u}}{a}\right)\left\{p + \frac{3P}{2d}\right\}^{-1} \tag{4.6}$$

but, at $r = a$, according to (4.4)

$$\bar{u} = A, \frac{d\bar{u}}{dr} = -qA \tag{4.7}$$

so that

$$A = \frac{aP}{dD}c_0\left[\left(p + \frac{3P}{ad}\right)\left(q + \frac{P}{dD} + \frac{1}{a}\right) - \frac{P}{dD}\frac{3P}{ad}\right]^{-1} \tag{4.8}$$

Before it is possible to recover the function $u(r, t)$ from its transform it is necessary to put this expression for A into a simpler form. It is convenient now to introduce the constant

$$x = \frac{aP}{dD} \tag{4.9}$$

This constant is a measure of how easily the material permeates through the surface. A small value of x, meaning that permeation is slow, results when P is so small compared with D that

$$\frac{P}{D} < \frac{d}{a} \ll 1. \tag{4.10}$$

For x larger than about $\frac{1}{2}$ the equations above would have to be modified since permeation would be so rapid that the concentration inside could no longer be assumed to be independent of position. In terms of x, A has the form

$$A = \frac{c_0 x}{D}\left[q^3 + \frac{1}{a}(1+x)q^2 + \frac{3x}{a^2}q + \frac{3x}{a^3}\right]^{-1} \tag{4.11}$$

The denominator is cubic in q and its roots could be found numerically but, for x in the range $0.1 > x > 0$, they are given sufficiently accurately by

$$\begin{aligned} \frac{q}{a} &= -1 - x, \\ \frac{q}{a} &= \pm i\sqrt{3x} \end{aligned} \tag{4.12}$$

This gives a factorization of the cubic which is correct except for a term of order x^2. Thus, to the same accuracy, A can be split up into partial fractions

$$A = \frac{c_0 x}{D}\frac{a^2}{1+5x+x^2}\left\{\frac{1}{q+(1+x)/a} - \frac{q}{q^2+3x/a^2} + \frac{(1+x)/a}{q^2+3x/a^2}\right\} \tag{4.13}$$

and the Laplace transform is given as

$$\bar{u} = \frac{a^2 c_0}{D}\frac{x}{1+5x+x^2}\left\{\frac{1}{q+(1+x)/a} - \frac{q}{q^2+3x/a^2} + \frac{(1+x)/a}{q^2+3x/a^2}\right\} e^{-q(r-a)} \tag{4.14}$$

To find the final solution for u it is necessary to know the functions for which each of these terms is the Laplace transform. The first term is of familiar form (Carslaw & Jaeger, 1953), but the others are not available in textbooks or are given incorrectly. For a derivation of the correct functions see the Appendix. Using these functions the final solution for the concentration outside can be written as

$$\begin{aligned} u(r,t) = \frac{a^2 c_0}{D}\frac{x}{1+5x+x^2}\Bigg[\frac{1+x}{a}\Bigg\{\sqrt{\frac{D}{\pi t}}\int_{r-a}^{\infty} e^{-s^2/4Dt}\cos(s-r+a)\sqrt{3x}/a\, ds \\ - D\exp[(1+x)(r-a)/a + Dt(1+x)^2/a^2]\operatorname{erfc}[(r-a)/2\sqrt{Dt}\Bigg\} \\ + (1+x)\sqrt{Dt}/a] + \frac{1}{a}\sqrt{\frac{3xD}{\pi t}}\int_{r-a}^{\infty} e^{-s^2/4Dt} \\ \sin(s-r+a)\sqrt{3x}/a\, ds \end{aligned} \tag{4.15}$$

or

$$f(r, t)/c_0 = a/r\, x/(1 + 5x + x^2)\Big[(1 + x)\{2/\sqrt{\pi}\int_{(r-a)/2\sqrt{Dt}}^{\infty} e^{-v^2} \cos \sqrt{3x}(2v\sqrt{Dt} - r + a)/a\, dv - \exp[(1 + x)(r - a)/a + Dt(1 + x)^2/a^2] \operatorname{erfc}[(r - a)/2\sqrt{Dt} + (1 + x)\sqrt{Dt}/a]\} + \sqrt{\frac{12x}{\pi}}\int_{(r-a)/2\sqrt{Dt}}^{\infty} e^{-v^2} \sin \sqrt{3x}(2v\sqrt{Dt} - r + a)/a\, dv\Big] \quad (4.16)$$

This general solution is of little value since tables of the indefinite integrals are not available.

5. The Concentration Inside and Just Outside the Sphere

It is fortunate that the general solution for the concentration outside the sphere is not required for most applications. The concentration just outside the sphere is of more interest and this is more easily calculated. When $r = a$ the general solution reduces to simpler functions and, in terms of the dimensionless variable

$$\tau = \frac{Dt}{a^2} \quad (5.1)$$

is

$$\frac{f(a, t)}{c_0} = \frac{x}{1 + 5x + x^2}\Big[(1 + x)\{e^{-3x\tau} - e^{(1+x)^2\tau} \operatorname{erfc}[(1 + x)\sqrt{\tau}]\} + \sqrt{\frac{12x}{\pi}}\, e^{-3x\tau}\int_0^{\sqrt{3x\tau}} e^{\beta^2}\, d\beta\Big] \quad (5.2)$$

The general behaviour of this solution agrees with expectations. As $t \to 0$ the first two terms tend to cancel while the third vanishes and as $t \to \infty$ the first term vanishes while the second and third tend to cancel. Thus $f \to 0$ at both limits. For small x and moderate values of t the first term is the largest.

As an indication of the nature of this solution for the concentration just outside, the expression above has been evaluated for $x = 1/10$ and $x = 1/60$ and the results are shown in Fig. 1. (A short table of the final integral is given in Jahnke & Emde (1945).) It is clear from these that the initial rise in concentration is very rapid and the fall comparatively slow. The maximum value attained is determined primarily by the first factor i.e.

$$f_{\max} \sim xc_0/(1 + 5x + x^2) \quad (5.3)$$

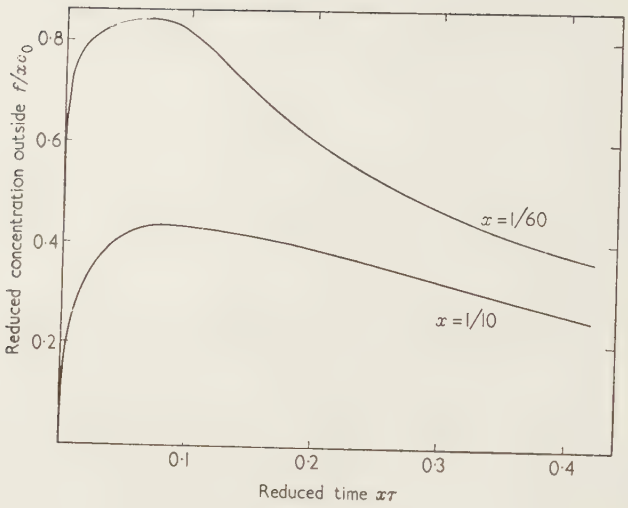


FIG. 1.

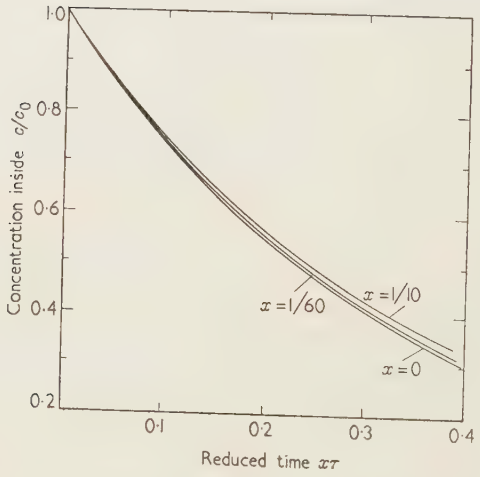


FIG. 2.

To have both curves in the same figure the variables are chosen as $\frac{f}{c_0 x}$ and $x\tau$. This, unfortunately, masks the fact that the initial rise is much more rapid for $x = 1/10$ and the maximum concentration much higher.

Once the concentration just outside is known, the concentration inside can be found from the equations of Section 3. In terms of the reduced

variables, the most convenient form is

$$c/c_0 = e^{-3x\tau}\{1 + 3x^2\int_0^\tau f(a,\sigma)/xc_0\, d\sigma\} \qquad (5.4)$$

The results of this are shown in Fig. 2. The integration was carried out numerically. The curves show clearly that, even for $x = 0.1$ the back-effect of the concentration outside makes little difference to the concentration inside except when both have dropped to small values.

6. Haemolysis from Erythrocytes

The particular example of haemolysing erythrocytes can be used to illustrate these calculations because the various constants are

$D = 6{\cdot}9 \times 10^{-7}$ cm^2/sec (Cohn & Edsall, 1943),
$P = 6{\cdot}9 \times 10^{-11}$ cm^2/sec (calculated from Heedman, 1957),
$d = 2 \times 10^{-6}$ cm (Ponder, 1958),
and $a = 3{\cdot}3 \times 10^{-4}$ cm

which are based on a normal volume of $90\mu^3$ (Albritton, 1952) and a volume increase of 70% on haemolysis (Ponder, 1948). These show that the permeation parameter x is approximately 1/60 so that the calculations above can be utilized immediately.

Figure 2, then, illustrates the fact that the decrease of haemoglobin concentration inside the cell is determined by the permeation constants of its membrane and is almost independent of the diffusion circumstances outside the cell. On the other hand, the escaped matter just outside the cell can rise to about $1\frac{1}{2}$% of the original concentration inside the cell (Fig. 1) and, when the cell is being observed by interferometry or absorptiometry, this may complicate the observations. The relation between the apparent profile of a cell in such experiments and its real profile is discussed elsewhere (Marsden & Hall, 1960).

The author is indebted to Dr. N. V. B. Marsden for suggesting this problem and for several discussions of its physical and biological significance. He is also grateful to Professor P. O. Löwdin, Quantum Chemistry Group, Uppsala, for his hospitality during this research and to King Gustaf VI Adolf's Seventy Years Fund for Swedish Culture, Knut and Alice Wallenburg's Foundation and the Swedish Natural Science Research Council for their generous financial support.

REFERENCES

Albritton, E. C. (1952). "Standard Values in Blood". Philadelphia.
Auer, P. L. & Murbach, E. W. (1954). *J. chem. Phys.* **22**, 1054.
Berthier, G. (1955). *J. Chim. phys.* **52**, 41.
Carslaw, H. S. & Jaeger, J. C. (1953). "Operational Methods in Applied Mathematics", 2nd Edition. Oxford University Press, London.
Cohn, E. J. & Edsall, J. T. (1943). "Proteins, Amino Acids and Peptides as Ions and Dipolar Ions". New York.

Heedman, P. A. (1957). *Exp. Cell Res.* **14**, 9.
Jahnke, E. & Emde, F. (1945). "Tables of Functions", 4th Edition. Dover, New York.
Marsden, N. V. B. & Hall, G. G. (1960). To be published.
Ponder, E. (1948). "Haemolysis and Related Phenomena". New York.
Ponder, E. (1958). *Protoplasmologia*, **10**, Part 2.

Appendix

The Laplace transforms required can be found as applications of the general theorem that the transform of

$$(\pi t)^{-\frac{1}{2}} \int_0^\infty e^{-u^2/4t}\, x(u)\, du \tag{A.1}$$

is

$$p^{-\frac{1}{2}}\, \bar{x}(\sqrt{p}) \tag{A.2}$$

where $\bar{x}\,(p)$ is the transform of $x(t)$. Thus the function whose transform is

$$e^{-a\sqrt{p}}/(p + b),\ a,\ b > 0 \tag{A.3}$$

corresponds to

$$x(t) = \begin{cases} \cos \sqrt{b}(t - a) & \text{if } t \geqslant a, \\ 0 & \text{if } t < a \end{cases} \tag{A.4}$$

and so is

$$(\pi t)^{-\frac{1}{2}} \int_a^\infty e^{-u^2/4t} \cos \sqrt{b}(u - a)\, du \tag{A.5}$$

By a similar method, or by differentiating (A.3) and (A.5) with respect to a, it can be shown that the function whose transform is

$$\sqrt{p}\ e^{-a\sqrt{p}}/(p + b),\ a,\ b,\ > 0 \tag{A.6}$$

must be

$$(\pi t)^{-\frac{1}{2}} \{e^{-a^2/4t} - \sqrt{b} \int_a^\infty e^{-u^2/4t} \sin \sqrt{b}(u - a)\, du\} \tag{A.7}$$

J. Theoret. Biol. (1961) **1**, 27–58

Quanta and the Concept of Organismic Law

Walter M. Elsasser†

Department of Physics, University of New Mexico, Albuquerque, U.S.A.

(*Received* 1 *September* 1960)

1. Introduction

A quarter of a century ago Niels Bohr first propounded a set of new and highly original ideas concerning the bearing of quantum mechanics on biological theory. They may be found elaborated in Bohr's later writings (1934; 1958). Unfortunately, there has been only a very limited follow-up of this work in the literature. This is usually attributed to the fact that the observational material is so complex that it does not lend itself to clear-cut conclusions and decisions. However, it is possible that, as Bohr already indicated, complexity itself, so prominent in all biological experience, is an essential ingredient of biological theory rather than just an inconvenient incident. In 1958 the author published a book oriented along this line of thought which, in spite of its preliminary character, was rather kindly received by the public. Encouraged by this, we have since been trying to clarify and develop many points: a presentation of some of the basic ideas in their current, more advanced state and without the expository material of the book, seems in order.

Our point of departure is that theoretical biology must profit from two main events in the more recent history of the physical sciences. One of them is the establishment of quantum mechanics as the theoretical basis for all molecular chemistry and physics, the other is the recent development of the theory of automata (although in this paper we shall not be much concerned with the latter). In the present investigation we start from quantum theory and try to penetrate from there toward certain generalizations that have a bearing on biology; only later in the analysis shall we discuss the relations of these ideas to results of empirical biology. The appearance of biological evidence at a relatively late stage might be objected to by some as being too speculative, but this would be misjudging the aims of our inquiry. Scientific reasoning in a still relatively obscure field must of necessity be inductive, excepting only well-defined special applications of established principles. Our arguments, even though they

† On leave of absence from the University of California, La Jolla, U.S.A.

start from general principles are essentially inductive. They lead from one set of abstractions to other more general abstractions. The very few working hypotheses which are necessary in this process of generalization must be exhibited as clearly as possible.

The view that biology is more general and not more special than physics and chemistry forms the basis of all our considerations; this view has been held by a great many of the leading biological thinkers throughout the history of the science (see Nordenskjöld, 1946). The distinction is a very clear-cut one; there is nothing ambiguous about it. As an example, it is well known that classical mechanics is a special case of quantum mechanics which can be obtained in the mathematical formalism by letting Planck's constant, h, tend to zero. Conversely, quantum mechanics may be obtained from classical mechanics only by superadding a basically novel and logically quite independent postulate of the formalism (non-commutativity of certain variables) with attendant changes in interpretation. In the historical development of the quantum theory there appeared first certain more or less arbitrary working hypotheses which led by very slow steps to the ultimate, correct theory. Again, it is not surprising that such a process of inductive generalization may lead first to the elaboration of some general principles and only in connection with those to a detailed comparison with experimental data.

It is our claim, then, that effective experimentation in molecular biology must be preceded or at least accompanied by analysis and, if necessary, inductive generalization of basic principles of physical dynamics and statistical mechanics, so as to exhibit just what can be stated in an operationally meaningful way when extreme complexity is an essential characteristic of the systems to be investigated.

This basic attitude leads us to a definition of the term mechanistic and its opposite, non-mechanistic, as used here. A mechanistic view is one in which it is assumed that biological phenomena can be described through specialization to very complex systems of the principles of physics, especially quantum mechanics, in the way this is generally accepted as true of all chemistry. The non-mechanistic view, on the other hand, assumes that significant generalizations are required before one can successfully penetrate from physics and chemistry into the dynamics of living systems. In the course of this inquiry we shall have occasion to uncover some of the implicit assumptions on which the mechanistic viewpoint rests.

2. Basic Concepts

At the present level our arguments are far more conceptual than mathematical. How far they may be converted into a more mathematical theory

is a decision which must be left to the future. Perhaps most biologists will not object to this state of affairs. It might not be an accident that the very complexity of their subject has kept them from indulging in too much mathematical abstraction, and for many problems this may be a distinct advantage. If the tools are conceptual, it is the more necessary that they be not only as precise as possible but also operationally meaningful and free from any metaphysical implications. After lengthy effort we have arrived at a basic vocabulary of five terms that seem to fulfil these conditions. These are:

Class
Law
Sample
Prediction
Complementarity

To these may be added a number of terms from the vocabulary of empirical biology and of standard physics which will not require a more thorough analysis in the present context. For example, "structure" is such a term. We shall now briefly discuss in sequence the terms named.

Class

A class is a set of objects that have certain properties in common. Beyond this, we need not here introduce a formal definition. It will also not be necessary to enter into any of the refinements of the mathematico-logical theory of classes. For our practical purposes it suffices to distinguish two types of classes, namely, concrete classes and abstract classes. The elements (or members) of concrete classes are definite objects of experience; the elements of abstract classes are themselves abstractions. Biological classes are concrete classes whose elements occur in direct experience. The class "cow" can be identified on a walk through the countryside. It is highly significant that this is not so for the most important concrete classes of physics and chemistry. Although we might be able to pick up a nearly perfect specimen of a quartz crystal on our walk, as a rule the most basic classes of inorganic matter are not so readily obtainable. Such a simple object as one gram of pure aluminum, or of pure fluorine, can only be had as the product of an extremely highly developed and complex scientific technology. Whereas almost all of the concrete classes of biology are given in more or less direct experience (at most by using a microscope) we would have to go back to the Aristotelian categories of air and earth, water and fire to achieve a similar immediacy in physical science. When it comes to such more refined concepts as the class of all hydrogen atoms it is very doubtful indeed whether it can even be defined as a concrete class;

one needs many years of experience in scientific reasoning to comprehend fully the implications of this concept.

Let us say at once that these ideas are only loosely connected with the old arguments of the philosophers regarding perception, reality, etc. Our aim is here to clarify the relationship of the basic concepts of atomic physics, which are *microscopic*, to the concrete objects of the biologist, which are *macroscopic*. In this paper the terms macroscopic and microscopic will be used in the sense of the quantum physicist, not in that of the empirical biologist: a macroscopic object is one large enough so that its main characteristics can be determined, in principle at least, without any appreciable disturbance of the motion of its elementary constituents such as is implied in any true measurement in the realm of quanta.

But to return to our class concept: the problem at hand is to match the concrete classes of biology with the fundamental classes brought to light by physical science. Now the latter are essentially homogeneous with respect to spatio-temporal and other physical characteristics (e.g. one gram of Al). When it comes to the investigation of energy or structure at any level of geometrical size the organism proves radically inhomogeneous everywhere, all the way down to the molecular level. Now practically the entire methodology of physical science rests on measurements made on homogeneous classes. Given the complications which occur in the concept of measurement when we go from classical physics to quantum physics, it is not at all clear how a system of such excessive structural complexity as an organism can be related to homogeneous classes. Let us say merely that the view of the mechanist rests on the assumption that such a relationship can be established in a reasonably straightforward way.

Law

A class is defined in terms of properties common to their elements; thus in the simplest case, we may define a class as the set of all objects (concrete objects or abstract propositions, as the case may be) that have property A (class A, for short). We speak of a law when all members of class A, or almost all members, also have property B. (The obvious case where B follows from A by mere logical implication may be disregarded here for simplicity). In this paper, we deliberately do not try to entangle ourselves in the complexities of mathematical rigor by requiring that B follow from A with necessity and without exception; thus we know that cows have four hooves, but we shall not raise any quarrel about a cow that has been mutilated and has only three. This kind of conceptual nonchalance will greatly simplify our business; it will permit us to cling to essentials without being too lengthy. There will be enough time to post a warning sign when this approach begins to lead us into trouble.

A *theory* is a set of laws mutually interrelated so as to form an abstract structure. We shall assume that the reader is sufficiently acquainted with the meaning of the last-named concept in modern scientific reasoning so that we need not enter into its discussion.†

Underlying any theory that is based on observed regularities are the tautological statements of the logical theory of classes which rest on a few simple definitions of abstract operations. The equivalence of two classes (class A = class B) needs no explanation. The concept of a "subclass" of a given class is also an obvious one. The "intersection" of two classes is the class made up of those members that are common to both the original classes. The "union" of two classes consists simply in the combined membership of both classes. The "complement" of class A is the set of all elements which have the property non-A. The formal calculus of classes (Boolean algebra) is based on these definitions and no others. It is really quite simple and can be learned with a little effort from a text on algebra or logic, e.g. Birkhoff & MacLane (1953). We have found, however, that there is no particular advantage in our using it explicitly in this discussion.

A theory, then, that is a system of laws which are not merely tautological, must state relationships that can be verified on concrete classes. The consequence of a theory, as of any single law, is to reduce the number of independent statements that can be made about certain types of observations. Any law states that one or the other of the aforementioned logical relationships exists between empirically defined concrete classes.

Sample

It is an essential feature of scientific methodology that not all elements of a given class can be investigated. We know that cows have four stomachs, but it would make little sense to investigate the truth of this statement for every cow. Thus it is an essential implication of the scientific method that a class can be represented with respect to observations by a sample of its membership. This constitutes a basic principle of inductive inference; it is commonly referred to as the "principle of the uniformity of nature". We need not argue about it but need to exhibit its character as a logically independent postulate. On the other hand, it will not be necessary for us to discuss the technical methodology of sampling as found in books on statistical analysis.

The principle of the uniformity of nature needs to be distinguished from a different property of certain classes, namely, *homogeneity*. We have already touched on the latter. To give an example, consider a piece of

† See Elsasser (1958), Chapter 5.

pure germanium metal. It is well known that the addition of a few ppm of certain impurities will radically change the electrical properties of this sample, converting it into an impurity semiconductor. Most inhomogeneities found in experimental problems are of a chemical type, but by no means all of them. Strain-hardening changes the characteristics of a metal by introducing purely physical, structural inhomogeneities. We can say, of course, that the metal with impurities or the strain-hardened metal represent different classes of objects as compared to the original, pure, and undistorted specimen. While this is true, we must still penetrate a little closer toward the exact meaning of inhomogeneity as it appears in the concrete classes of physics and chemistry.

The corresponding theoretical concepts are developed in statistical mechanics. One starts out with a system having N atoms or molecules and hence αN dynamical variables, where N is an extremely large number and α is some number not very large compared with unity (for instance, $\alpha = 6$ in the case of a monatomic gas). By means of processes of statistical averaging all but a few, say β, of the dynamical variables are eliminated and it is claimed that the macroscopic properties of the substance can be described satisfactorily in terms of these β variables while all the other variables may be ignored. On modifying a system slightly, one may change the values of these variables and sometimes also increase or decrease their number (e.g. adding impurities to Ge).

A little inspection now shows that the term homogeneous is used in two senses: First, we call a class more homogeneous than another class if β is smaller for the first class than for the second. Next, we call a class homogeneous if the variation of the number β of its parameters from one element of the class to the next is small. Homogeneity, therefore, refers to the methodology of preparing or selecting classes. This process will in general involve inductive inference and will depend on the laws holding for the classes considered; but we need not assume, in physical science, that the validity of the laws themselves is being questioned.

Prediction

Here, we can be brief. Consider a general logical statement, a "predication", for instance: if A and B, then C. This has no specific reference to time. We speak of a prediction if C corresponds to a later time than A and B. In practice almost any scientific predication can be converted into a prediction, and this is done as a matter of course since all observation and experiment takes time. Thus in place of the proposition stated above, we say that if A and B are fulfilled, an observation will show C. Speaking of prediction has the advantage of obviating the use of the concept of causality. It is clear that causality is equivalent to the possibility of making

predictions. The writer has found, however, that the scientific public, including some very distinguished men, are not in agreement about the precise meaning of the term causality in biology, and he prefers to avoid it.

Any law and any theory may give rise to prediction. Some students of paleontology and evolution have tried to speak of "non-predictive" theories. It seems to us that this is inadmissible; such a "theory" would be merely a description or classification. A theory is a system of relationships possessing abstract structure. It should at least enable us to say that if you find A and B fulfilled in your next paleontological discovery, you will also find C. Beyond the realm of such statements there exists no theory in our definition.

Complementarity

We first define the concept of a measurement in terms of classes. One calls a set of classes *disjoint* if no two of them have an element in common. We define a measurement on an object as the assignment of the object to one of a set of disjoint classes. As an example, consider the measurement of a length x. We divide the whole range over which x can vary into a series of intervals $\Delta x_1 \ldots \Delta x_i \ldots$ which are non-overlapping and cover the entire range. The measurement assigns the value which x has in the given object to one interval, say Δx_i. We can of course make the mesh of intervals as fine as we please, so this definition in terms of disjoint classes does not imply any limit on the precision of measurement.

The definition of a prediction is somewhat more involved. It is not in general possible to make a unique assignment in advance to one of a set of disjoint classes; instead, we can only assign a numerical *probability* to each class. Thus we assign the probability p_i to the interval Δx_i. This implies that the prediction does not refer to a single object but to a concrete class. The probabilities represent frequencies within classes, in keeping with the usual frequency interpretation of the probability concept. (For formulas, see Appendix.)

We now consider pairs of complementary predictions as they appear in the wave-particle duality of quantum mechanics (Bohm, 1951). This duality can be expressed by saying that if we try to determine at the same time the position and the velocity (momentum) of a particle, say an electron, we can do so only by admitting a certain *spread* in either (or both) of these "complementary" quantities. Quantum theory tells us that if we determine the position of the electron with a small spread, then the spread of the prediction we can make about its velocity at the moment of measurement is necessarily large, and conversely.

We can now readily generalize the concept of complementarity: Suppos we want to predict a quantity, say A (A may stand also for a set of quan tities) and we can assign mathematically a certain spread to this prediction Next assume that we want to predict another quantity, B (or set o quantities), characterized by another spread. We say that the prediction of A and of B stand in a relation of complementarity if there exists mathematical relationship between the two spreads of such a nature tha an increase in the first spread implies a decrease in the second, and con versely. This definition is quite general, as general as our definition o measurement. It should be said at once, however, that we do not envisag complementarities that are novel compared to the basic wave-particl duality of quantum mechanics; we consider all complementarity founc in nature as representing consequences directly deducible from this basi dualism. Merely, in complex systems one often deals with quantities tha are very different from positions and velocities of the electrons anc nuclei, although in principle the theoretical physicist will be able to express any measurable quantity as a mathematical function of the last-named ones. If many elementary particles and their interactions are involved, complementarity may assume rather complex forms, but it seems certain that any such form can be subsumed under the preceding general scheme.

Parenthetically, we need to supplement the preceding discussion because so far we have spoken only of pairs of predictions as distinct from pairs of measurements. Now if we admit measurements with a finite error distribution, we may think of one of the two complementary operations as a measurement and of the other as a prediction. We cannot think of both as measurements because after having measured A, say, there results a certain minimum spread in the prediction of B; this spread in B may be arbitrarily reduced by a subsequent measurement, but only at the expense of widening again the corresponding spread in A. The preceding arguments are put into more precise, mathematical language in the Appendix.

3. The Import of Quantum Theory

We shall now ask ourselves to what extent quantum mechanics is valid in the living organism. By quantum mechanics we mean that part of the theory which applies to the energy range encountered in conventional organic reactions (remaining in practice below, say 30 eV or 700 kcal). In this range of energies, nuclear effects and the effects described by relativistic field theories give rise only to minute correction terms, as a rule negligible. The author, himself a theoretical physicist, believes it the consensus of his colleagues that future progress of knowledge in the high

energy field can have no significant bearing on the type of quantum mechanics applicable to organisms, just as progress in atomic physics had no bearing on the older laws of the motion of celestial bodies. Thus if we speak of deviations from quantum mechanics we can only have in mind specific modifications of the theory related directly to biological phenomena. There is no empirical evidence whatever that such modifications are needed. Instead, the distinguishing feature of biological systems is their enormous structural complexity. It would be strange indeed (although no such possibility can of course be refuted *a priori*) if mere complexity gave rise to a breakdown of the principle of the uniformity of nature. On maintaining this principle, we shall postulate that the laws of quantum mechanics hold without exception in the organism and are invariably found valid whenever they are subject to experimental test.

Will this assumption restrict us to a purely mechanistic biology? The appearances are indeed so as we shall presently see. We shall deal with this problem by an investigation into the methodology of physical science which brings out its limitations when applied to systems as enormously complex as organisms. At present, let us agree for the sake of the argument that organismic laws are broader and not narrower than the laws of physics. Now within the universe of discourse established in the preceding section this can have only one meaning, namely that the laws of organisms have components that are not deducible from the principles of quantum mechanics. Such components will be designated as *biotonic*. Leaving the detailed explanation of this term for later, we merely assume the existence of these components and proceed to show the limitations imposed upon them by quantum mechanics.

These limitations may be expressed in terms of what we shall call the *theorem of means*, a mathematical proof of which is given in the Appendix. Here we indicate the argument in outline: Given a concrete class of physical objects, assume that we measure some physical quantity on each element of the class. One designates as the expectation value of that quantity the linear average, or mean, of the measured value taken over all elements of the class. The expectation value for a merely predicted quantity is defined in a corresponding way; that is, although the individual predicted values may have a probability distribution, the expectation value will again be defined as their mean over the class. Now assume that the class itself is defined in terms of certain parameters which have the same values at initial time, $t = 0$, for all members of the class. (Here, $t = 0$ does not of course mean the same time for all specimens relative to an external clock; it means an initial time chosen for each specimen so that it conforms to certain class characteristics at that moment.) Quantum mechanics permits us then to predict the expectation value or mean over

the class for any physical quantity whatever at any later time. The theory says that these predictions are uniquely determined after the manner of Newtonian mechanics. In other words, while prediction in quantum mechanical systems is intrinsically statistical for the individual system, the expectation values, which are class averages, are uniquely defined for all times. This is true for a closed system; we shall discuss this particular limitation later. This theorem holds for any physical quantity, whether macroscopic or microscopic. Among these physical quantities is also the standard deviation of any statistical distribution as well as all its higher statistical moments. These determine the probability distributions, all of which are therefore also uniquely determined if taken over the entire class.

Of course, we can make a microscopic, that is, individual molecular or atomic measurement only at the expense of disturbing the system severely. Such a system can thereafter no longer serve as a sample of the original class, as has been emphasized by Bohr (1934; 1958). We must therefore restrict the measurements of our initial parameters as well as those later measurements which do not eliminate the sample from the class, to types that may be designated as non-perturbing, that is, they must be sufficiently far above the domain of quantum measurements with their inevitable perturbations so as to leave the measured system reasonably intact (Elsasser, 1937). We think that these notions are clear enough that we do not have to resort here to detailed quantitative definitions.

The theorem of means is extremely powerful. It tells us that the expectation values of all physical quantities are determined for all times as function of the initial conditions, in any closed system. As a corollary, if biotonic components of organismic laws exist, their effects must average out to zero for the expectation value of any physical quantity whatsoever. It is hardly necessary to enlarge on the extreme stringency of this requirement and on the hopelessness of fitting biotonic regularities into a realm of abstract relationships where all their class-averages must vanish. If, therefore, we wish to remain within that tradition of biology which claims that organismic laws are broader and not narrower than the laws of physics and chemistry, we must somehow show that the validity of the theorem undergoes suitable restrictions in organisms. Given the applicability of quantum mechanics which we postulated, this can only mean that organisms are specifically so designed as to limit the experimental verification of the theorem of means. It is the principal aim of this paper to investigate the possibilities which exist for such design.

There are two main conditions in the world of organisms which allow them to exhibit biotonic characteristics notwithstanding the theorem of means. One of these is readily apparent and has been widely discussed in

the literature; this is the fact that all organisms are open systems. It is impossible to doubt that this fact plays a determining role in the dynamics of organisms. From the viewpoint of statistical mechanics the most significant characteristic of a closed system is that once a steady state is achieved this state is necessarily one of thermodynamical equilibrium, and then the principle of detailed balancing holds. This principle imposes the most severe restrictions on the dynamical configurations possible in equilibria; and upon their relaxation in open systems there is undoubtedly a tremendous increase in the variety of dynamic processes of which organisms can make use.

On the other hand, the fact that organisms are open systems does not in itself allow one to resolve the difficulties of the theorem of means. Consider an open system which is in a steady state, matter as well as energy coming in at a uniform rate and leaving at the same rate. In this case the mathematical theory of the open system is extremely similar to that of a closed system, except that in an open system the margins of fluctuation of many physical quantities will be very much larger than in a closed one. On the other hand, there is enough evidence, considering the general structure of the theory, that such generalizations of the theorem of means as may be derived for open systems are only a little less restrictive than those for closed ones. On superficial view there seems to be the difference that in an open system the constituents change constantly. In quantum mechanics, however, the theory does not permit us to distinguish between atoms of identical species in any event, and there are ready-made mathematical tools for going from the closed to the open system; the distinction between the two types of system is not nearly as radical as one might be inclined to think.

All this requires implicitly that the input of the open system be reasonably homogeneous (in the exact sense defined above, where the number of parameters characterizing the input is relatively small). Now many lower organisms and most plants can be grown from inorganic nutrient solutions of a relatively simple constitution, simple at least compared with the tremendous structural complexity of organic tissue itself; hence we may say that their input is reasonably homogeneous. We may draw the conclusion that while open systems offer many more possibilities for the dynamics of organisms than closed ones, there seems to be a need for a broader and more penetrating principle in order to elucidate the relationship of living matter to the theorem of means.

Such a principle has been outlined in a rather qualitative fashion in the book quoted above (Elsasser, 1958, Chapter 4). To appreciate its significance, let us first say what we ought to expect of it. It must express design principles embodied in the organism, or more generally speaking

in the world of organisms, which limit the validity of the theorem c means. These principles cannot have the character of formal laws on th same level of rigor as the laws of physics, because then we would have t admit that the laws of quantum mechanics are replaced by different law and hence do not hold in their usual form in the organism, in contradictio to our basic assumption.

At this point, we return to the earlier remark that complexity itself ma be an essential rather than an incidental property of organisms. The basi requirements in the conventional methodology of physical science ar twofold: It is implicitly assumed that classes are homogeneous, or can b made homogeneous by suitable selection of subclasses; next, it is implicitl assumed that the membership of classes is potentially infinite. Whenev it is found that within a class there are large variations of properties, w divide this class into subclasses and continue doing so until we hav arrived at classes that satisfy us as being sufficiently homogeneous for ou purposes. This again implies that we do not run out of specimens in th course of this separation into subclasses, it being assumed that any sub class we wish to prepare still has an arbitrarily large number of member of which we could avail ourselves if we wanted to.

We now introduce the following working hypothesis: *The assumption of the homogeneity and of the potentially infinite membership of classes do not apply to biological classes.* We consider this the essential distinction between the methodology of physical science and that of biological science This makes sense only if the number of specimens of any biological class available for investigation is actually finite and bounded, and we shall of course have to discuss this particular aspect of our hypothesis. We note that on this assumption complexity does become an essential ingredient of the theory. Organisms are structurally and dynamically so complex that one can always find individual differences, at least microscopic ones in the sense of our use of the term, between any two organisms of the same class, no matter how one has defined the class.† Thus on sufficiently close investigation the homogeneity of any class of organisms will disappear. This is certainly not a logical feature of any known theory of physical science. We have called this *the principle of finite classes* (Elsasser, 1958, Chapter 4).

Even before entering into a discussion of its precise meaning and implications, we recognize that this principle gives to the theory of organisms a different logical structure from conventional physical theory. Limiting processes carried out in terms of operations on an infinite sequence of members of a class are no longer meaningful. If the existence of such

† Ideas of this kind can be traced in a rudimentary form to Descartes and Pascal.

sequences within homogeneous classes is considered the characteristic of physical law then, clearly, the biotonic components of organismic law do not have this characteristic. There can then be questions which are essentially undecidable: before one is able to construct a class of sufficient homogeneity so that the question could be operationally decided by measurement, one runs out of specimens. In view of the stringency of the theorem of means, we shall postulate that all biotonic components of organismic laws represent regularities within finite classes, of such a nature that their rigorous deducibility from the laws of quantum mechanics can be neither established nor disproved by direct experimentation on the elements of such a class.

It is important, however, to note that this last statement is not a new postulate. It merely restates the fact that the methodology of homogeneous classes of virtually infinite membership will break down in the course of any attempt to relate biotonic regularities to pure physics and chemistry by the conventional procedures of precise experimentation as understood in physical science. This does not mean that biotonic regularities or laws cannot be verified. (Biotonic "laws" are now of course meant to indicate regularities applying to finite classes of internally inhomogeneous objects, as distinguished from physical laws which are implicitly assumed to apply to homogeneous classes of virtually infinite membership.) We observe biotonic laws in the first place as properties of the concrete classes of biology. If we now try to decide experimentally whether or not these properties can be derived deductively from the laws of quantum mechanics, we are prevented from finding an experimental yes-or-no decision by the inhomogeneity and finiteness of the concrete classes of biology. Instead, the relationship of biotonic regularities to the laws of physics must be established inductively. Now inductive inferences, while they never reach the rigor of formalized deductive arguments, can still be made convincing by way of progressive accumulation of evidence. In the realm of logical induction this process of accumulation replaces isolated crucial experiments. This brings us back to what we said before about the central role that design features of organisms play in this type of biological theory. Combining Bohr's conclusion that one cannot exhaustively determine the microscopic structure of a system without destroying it, by virtue of the complementarity relations of quantum mechanics—combining this with the internal inhomogeneity of all concrete classes of organisms, we are deprived of the possibility of straightforward and unique experimental decisions in the realm of biotonic phenomena. Instead, we can study the design of organisms and their operations (that is, we take the word design in a dynamical rather than a purely static sense) and by this study we are led to inductive inferences about the nature of biotonic laws and their

relationship to the basic laws of quantum mechanics. As we proceed, w hope to show that this approach can indeed lead to significant interpreta tions of the relationship between biological phenomena and the laws c physics and chemistry.

4. Microscopic Structure and Dynamics

We proceed now to analyse more closely the idea that classes of biolog differ from those of ordinary physical science by their inhomogeneity an the finite size of their membership. In statistical mechanics, a well-define microscopic configuration of a system (in phase space) has been designate since Boltzmann as a *complexion*. Disregarding velocity distributions, w may explain this term by the model of a purely geometrical, structura arrangement: given a crystal of a binary alloy, say, whose two component intercrystallize freely, for instance AgAu. In each lattice site there may b either a silver atom or a gold atom. Each complexion is represented by distribution where a certain number of sites are occupied by Ag, th remainder by Au. Roughly, leaving all mathematical refinements asid the number of possible complexions is of order

$$Z = N^N$$

where N is the number of atoms. To fix our ideas, consider a grain o this material comparable in size to a typical cell; for instance, let $N =$ 10^{14}. Since Z is a very large number it is convenient to use its logarithm

$$\log Z = N \log N$$

which, in our example, gives $\log Z = 1{\cdot}4 \cdot 10^{15}$ (using logarithms to bas 10). This number is also extremely large, of the same general order as N itself. (Those familiar with statistical mechanics will recognize in $\log Z$ a measure of the entropy, in our particular case of the entropy of mixing. Numbers of such tremendous magnitude that their logarithms are them selves very large numbers will be referred to as *immense*. The preceding result can be generalized: the number of complexions, that is, of differen possible microscopic arrangements of elementary constituents (whethe they be atoms, radicals, or small molecules) is an immense number in the sense defined. Now the living cell has, apart from its water, a quasi- solid structure, and if we admit that at least the weaker chemical bonds can be broken and rearranged, the number of complexions possible for a cell is also immense, of general order N^N, where N may be taken to be the number of atoms, radicals, or small molecules whose internal bonds are assumed to remain fixed in the course of the rearrangements admitted.

Next, let us try to obtain an estimate of the total number of cells of a given class which an idealized observer might be able to procure. We do this by considering the total number of cells on earth as an upper bound

to this number. Again, we are not interested in cells as such but in their microscopic complexions which will change rapidly; hence we propose to count cell-complexions, assuming that the complexion changes substantially, say, every minute of time. Previously (Elsasser, 1958) we calculated on assuming a maximum of 10^9 cells per cm^2 of the earth's surface, and a time span of 15 billion years or 10^{16} minutes, that an upper limit to the total number, say ν, of cell-complexions is about $\nu = 10^{44}$. If we assume that there is life on 10^{20} other planets, this number becomes $\nu = 10^{64}$. It is not our purpose, however, to indulge in cosmological speculations! The physical implications of our argument are contained in the inequality

$$\log \nu \ll \log Z$$

which says that the number of cell-complexions that can be available to even the most idealized observer is extremely small compared with the total number of possible cell-complexions which could be produced *in abstracto* by steric and dynamical rearrangements. The cosmological speculation merely assists us in visualizing the physical meaning of the principle of finite classes, which latter represents essentially a working hypothesis and not part of a speculative set of cosmological ideas. (In classical mechanics, the Euclidean character of space is a working hypothesis which can be used in practice very successfully without any reference to the cosmological problems that appear when the assumption of Euclidean space is pushed to its very limits.)

Nothing about the immense number of complexions could be said to pertain to biology proper. All the preceding is standard knowledge in the statistical mechanics of homogeneous inorganic systems. Thus if we speak of complexions as being significant for biology, we must imply that their role in biology differs from that in conventional inorganic science. Ordinarily, in statistical mechanics, the microscopic variables describing the nature of the individual complexion (the number of these variables being of order N) are averaged out and only a small number, β, of variables remain to describe the macroscopic behavior of the system. But organisms are so tremendously complex that they cannot be so described, or rather, only for a very short span of time. Their complexity, or inhomogeneity, prevents us from singling out a small number of parameters in terms of which the dynamics of the system could be represented in the long run.

Now, as a matter of fact, if we look at organisms purely statically, their most significant constituents (protein molecules and enzyme systems, chromosomes, ribosomes) are perhaps not as inhomogeneous as our argument might suggest. But experience and theory (Born, 1958) show equally that even moderate inhomogeneities do in the long run have a profound influence upon the dynamics of a physical system. It is therefore appro-

priate to think that the basic inhomogeneity of organic systems exerts its influence dynamically, in the course of the complex metabolic processes that characterize the living cell. We might say that the biological implications of microscopic inhomogeneity involve the time; they can only be understood four-dimensionally, in a customary manner of speaking of the physicist, not purely statically and three-dimensionally.

At this point we should emphasize again that the organism manufactures its own structural inhomogeneity; it does not import it from the environment. Misconceptions have sometimes arisen here through generalization of somewhat vague biological arguments. It is true that the development of the nervous organization of many higher animals depends critically on proper interaction with their social environment. But in our context we are dealing with the complexity of the basic physico-chemical structure of the cell and of its metabolism. It is well known that bacteria and plants, for instance, can be grown from nutrient solutions that contain little more than a few simple salts (plus a source of carbon in the case of bacteria). Even in higher organisms the proteins ingested must be broken down into very simple polypeptides before they can be utilized. Thus, the tremendous structural and dynamical complexity of the living cell is produced by the cell itself; it is endogenous. It seems to us that this fundamental fact has not always been sufficiently stressed in the biological literature.

We now come to the application of these principles to the concept of organismic law. How do these circumstances open up the way for organismic laws that are broader than, and not confined to, the consequences of quantum mechanics? We proceed here in two steps. The first is the introduction of Bohr's (1934) concept of generalized complementarity. In the microscopic realm (realm of quantum mechanics) any measurement, taking necessarily the form of a physical interaction with the measuring device, engenders a perturbation whose exact magnitude cannot be determined in advance (only the average over many cases can be so determined). We need not confine ourselves to measurements of particle positions which communicate extra velocities to the particles; any number of other dynamical variables might be measured (for instance one might try to localize, at a given instant, the energy of a set of resonating homopolar bonds). In this as in any other case of microscopic measurement the interaction between the measured object and the measuring device limits the accuracy with which the microscopic state of the measured object can be determined after the measurement, and hence limits the accuracy of prediction of the future behavior of the object. Mathematical analysis shows (Born, 1958) that even in classical physics an indeterminacy in one dynamical variable spreads rapidly over all the variables of the system, thus making prediction progressively more inaccurate as time goes on;

the same applies in quantum theory. We conclude with Bohr that there exists a relation of complementarity as between our knowledge of the microscopic structure (complexion) of the system and the prediction of its undisturbed behavior as time goes on. If we make elaborate measurements, precise enough to determine the microscopic state of the system at a given instant, we can indeed find out what this state is, but the disturbances engendered (for instance breaking of chemical bonds) would be so radical that the system would behave thereafter in a quite different way from the way it did before; it can no longer be considered as the same dynamical system. In practice, as Bohr says, we have killed the organism by our too detailed measurements. Conversely, if we want to conserve the system we must confine our measurements to sufficiently non-perturbing ones; but this implies that the mesh of our measurements is so coarse that we cannot determine a precise microscopic complexion; we can merely limit the number of complexions that may be realized in the system, but this remaining number of nearly equally possible complexions will still be immense. Thus, prediction of the undisturbed behavior of the system is complementary to a knowledge of its microscopic complexion, in the sense in which we have defined complementarity before. The quantitative aspects of this complementarity are open to analysis by methods of quantum mechanics.

The next step in our analysis consists in taking account of the fact that organisms appear in classes. This leads to the presumption, which we can hardly avoid, that by virtue of the principle of the uniformity of nature we can sample a concrete class and restrict our measurements to the samples. We may then assume that we can determine the microscopic state of each sample precisely at the expense, however, of destroying the sample so that we must discard it thereafter. The question is whether this type of measurement applied to an appreciable numerical fraction of any one class can help us to make statements about the remaining, undisturbed membership of the class. It is clear that in this way we can gather a large amount of information about the class, in so far as the class is homogeneous, and we may admit that in this way we can procure all knowledge about the class that is compatible with the assumption of its homogeneity. But this homogeneity, while far-reaching in macroscopic features, is limited in the microscopic realm by the principle of finite classes as expressed in the inequality given above. This inequality, derived originally for any physical system containing a very large number, N, of atoms (or radicals and small molecules) may now be interpreted in straightforward biological terms: a concrete class of biology is characterized by a set of macroscopic parameters. The principle of finite classes says that the number of microscopic complexions compatible with the class characteristics is immensely

large compared with the membership of the class. Hence the precise determination of any sampling of the class (even if we used almost the entire class as samples) leads only to the determination of an immensely small fraction of all complexions that are possible for the class. Thus the sampling of any fraction of the class cannot yield a significant prediction about the complexions realized in the remainder of the class. Thus the principle of finite classes supplements Bohr's principle of generalized complementarity in showing that predictions about the future dynamical behavior of organisms are limited not only for the individual organism but also for the class.

It is readily seen that the pursuit of this line of thought may lead us into somewhat intricate problems of epistemology. We have dealt with these very briefly elsewhere (Elsasser, 1952) and this paper is not the place to wax too philosophical. Certain aspects of these problems are operationally formulated within the framework of the conventional interpretation of quantum theory (Bohm, 1951), and the pertinent statements can be made without danger of our losing ourselves in vague philosophical speculation. Anybody who has even a slight familiarity with the principles of quantum theory has encountered the following basic example: Given an isolated electron, it is not in general possible to state with certainty the direction of its spin. Instead, we can only assign probabilities to different spin directions (which probabilities may become certainties in limiting cases). The values of these probabilities depend essentially on the antecedent interaction of the system, that is, on its physical history. This is a universal trait of systems of quantum mechanics: in general, we can only assign probabilities to microscopic variables, and the magnitudes of these probabilities depend again critically on the antecedent history of the system. This might be restated by saying that the only statements we can make about the microstructure of a system are inductive-probabilistic ones.† The term inductive has here its usual logical meaning, referring to inductive inference. These things are well known in the ordinary applications of quantum theory to homogeneous systems, and such conceptual or mathematical difficulties as appear in these cases have largely been mastered. But organisms are vastly more complex. Since organisms are highly inhomogeneous and also constantly manufacture their own internal complexity, the problem of inductively determining probabilities for the immense numbers of possible microscopic complexions becomes

† We wish to point out that the division of indeterminacies into quantum-mechanical ones and those pertaining to conventional statistical mechanics, which latter can be removed by reducing the system to a "pure state" (as used by von Neumann, 1933) is merely a mathematical artifice not applicable to the description of concrete physical classes (see Elsasser, 1937).

quite difficult. This inductive assignment of probabilities has to be carried out, of course, in function of the known characteristic of the concrete class at hand. There exists no unique prescription or recipe of how to evaluate these inductive probabilities; they do depend critically on the specific measurements which are possible with samples of the class.

We shall next return to our primary assumption, namely, that organismic laws have biotonic components which cannot be derived from physical laws. We are now able to give a more precise meaning to this statement. As pointed out previously (Elsasser, 1958), there is only one kind of measurement but there can be different kinds of prediction. A measurement is a physical act, an interaction of the measured object with a measuring device, in the course of which there is left a definite physical trace in the measuring device (at least it seems quite safe to objectify all meaningful measurements in this way). A prediction on the other hand is purely physical only if it is based on the solution of the differential equations of physics (quantum mechanics) with reference to some future time while utilizing the results of the available measurements (data) referring to the past and present of the system (or class). The concept of organismic laws which are broader and not narrower than physical laws is operationally meaningful only if these laws imply predictions that are not in their entirety physical predictions as just defined. If such more extensive predictions are possible we say that the organismic laws have biotonic components. (We do not claim that biotonic phenomena can ever exist "by themselves", that is, dissociated from a framework of physical structure, which latter will always permit a minimum of physical prediction.) We have seen above how the stringent requirements of the quantum-mechanical theorem of means may be overcome by virtue of the principle of finite classes. We can now formulate this as a necessary condition for the possibility of biotonic components of organismic laws. Predictions involve biotonic components if they cannot in their entirety be reduced to purely physical prediction. The biotonic regularity leads to the prediction that certain events occur in the members of a biological class with high probability. This must not contradict any possible physical prediction for the class, which in turn implies that the probabilities resulting from any physical prediction must in all cases have a sufficiently low measure of confidence.

In order to make this necessary condition a sufficient one it would be required to construct a model of organismic behavior based on the mathematical theory of inductive probabilities; such a model would have to demonstrate more rigorously the mathematical possibility of biotonic behavior within a finite class endowed with sufficient physical indeterminacy and low confidence of physical prediction. Such an endeavor is

clearly still very far in the future. All we can try to do here is to indicate the existence of these problems. We shall proceed somewhat farther in a more concrete direction, by endeavoring to relate the preceding abstract ideas to observed characteristics of organisms.

5. The Theory of Epigenesis

What is the relationship of the concepts so far discussed to existing theories of biology? Many such theories are of course simply applications of known principles of physics and chemistry. Among the more purely biological theories of broad application is the Mendelian scheme of genetics, but this is so strongly empirical that it does not seem to lend itself readily to a verification of the ideas discussed here. Next, we shall avoid the introduction of any reference to the theory of evolution because we believe it is adequate, within our context, to deal with a universe of stationary biological classes, corresponding to a pre-evolutionary viewpoint. Again, among the remaining general theoretical concepts of biology, *epigenesis* appears pre-eminent owing to its broad fundamental significance as well as in view of its remarkable history (appearing first in the early eighteenth century when the controversy between epigenesis and preformation arose, the term is still in current use in the literature). This author at least feels that in dealing with epigenesis one is on ground that has been fertilized by and is germane to the mode of approach of experimental biology, especially morphodynamics. As is known, the theory of *preformation* claims that all the structures of the adult organism are specifically contained in the germ cell and "unfold" in the process of embryonic development. The theory of epigenesis claims in its traditional form that these structures are contained in the germ cell merely as "potentialities" and that they come into being through processes that are organismic. The precise meaning of these terms in more modern language will now have to be determined.

We shall here introduce the concept of "information". Since it is admittedly vague we use it merely as an intermediary tool to exhibit the ideas, and shall soon try to recast the language into more precise form. Let us also for the moment assume for the sake of simplicity that information comes in distinct parcels. We may then use the language of the mathematician when he deals with mathematical representations (mappings) (Birkhoff & MacLane, 1953). Three cases are shown schematically in Fig. 1 where each dot is meant to symbolize a parcel of information; (a) represents a one-to-one correspondence, designated as an *isomorphism*, (b) represents a one-to-many correspondence, and (c) a many-to-many correspondence; the latter two are designated as general *homomorphisms*. In each case we have labelled the upper line by G (to remind us of genes)

and the lower line by M (to remind us of morphology), but the scheme itself is, of course, quite general.

We shall now tentatively identify scheme (a) with the preformationist viewpoint and scheme (b) or preferably (c) with the epigenetic viewpoint. The essential point of the latter schemes is (using again our loose terminology) that in (a) the line M contains the same or an equivalent "amount of information" as line G, whereas in (b) and (c) line M contains "more information" than G. It is clearly necessary to explain what this means. To speak of "information" in biology, we must deal with the structure of

FIG. 1.

classes of organisms. Much of this structure will be macroscopic, but our main requisite is now, not so much that it be macroscopic but that the existence of structures as class properties can be ascertained by an examination of a suitable sampling of the class. Now biological structure is variegated and awkward to deal with. To overcome this we introduce an artifice: we assume that it is possible to obtain a description of the class structure. By this we mean a message which is a sequence of symbols. We assume that the symbols used in the description are defined by means of a dictionary. In the case of biological systems the dictionary may contain geometrical, mechanical, chemical, optical, and other terms. Using such a dictionary an observer can translate the measured class characteristics into a description. We shall not specify here the size of the mesh of observations which characterizes the dictionary and thereby the accuracy of the description; we leave this open, depending on the particular aims of the observer, provided only it remains the same in the course of time for a given series of observations.

Now consider a class of organisms that undergo, say, embryonic development. Keeping the dictionary, the technique of description fixed, we see that the message, the description itself, changes with time; the message undergoes a transformation. We are thus led to study the transformation of messages, which is the subject of communication (information) theory and also of the theory of automata. Without entering into these subjects which the author has discussed elsewhere (Elsasser, 1958), with a view to biology, the concept of the transformation of a description will help us to clarify the relationship of the theories of preformation and epigenesis.

Consider first the theory of preformation. As Fig. 1 (a) schematically indicates, one assumes it possible to find a description in which the transformation of the information content of the class in the course of embryonic development, say, is a one-to-one correspondence, an isomorphism. It cannot of course be claimed that this is true of any possible description, but one assumes that there are some descriptions, or at least one for which this applies. Now this kind of transformation is ideally exemplified by *classical physics* where there are no limitations, in principle, to the accuracy of measurement. The dynamics of a physical system is expressed by means of a set of differential equations. The integration of these equations as a function of time may be interpreted mathematically as a transformation of the description of the system, namely, from the description at an earlier to that at a later time. We can express this conveniently in terms of prediction, one of the basic concepts of our list in Section 2. The prediction of such a system is fully determined by the knowledge of the parameters of the system at the initial time (Newtonian type of causality). This implies that the transformation is general, by which we mean that it involves only abstract operational rules (namely, of how to integrate differential equations) which contain no reference to the specific system under consideration. Again, in microscopically more complicated systems which must be described by quantum mechanics, this represents the case of homogeneous classes where there is a neat distinction between macroscopic and microscopic variables and the former are connected with each other by differential equations, thus admitting of unique description and prediction. If we apply this to the type of description of biological classes assumed by the theory of preformation, we may well say that this theory embodies what we have called the mechanistic view of biology. Whether or not the biological classes considered are actually homogeneous is not so much the question as whether their description can be carried out along a pattern akin to a macroscopic description of homogeneous systems. Since in this case the transformations which connect successive descriptions are straight isomorphisms it is in general correct to assume that they have an inverse.

This means that if in Fig. 1 (a) there exists an isomorphism carrying G into M, there also exists the inverse isomorphism carrying M into G. In concrete, biological terms this implies that on the theory of preformation the germ cell must contain an isomorphic mapping of the structure of the corresponding adult organism (and the same would no doubt apply to many of the non-degenerate somatic cells).

We now turn to the theory of epigenesis. We shall schematically represent the transformations occurring in the process of embryonic growth by Fig. 1 (c). In our qualitative language this would mean that the "information content" of the class "increases" in the course of embryonic development. There is no point in saying that there exists no transformation which carries G into M, since we can always define, by brute force as it were, such a transformation. We can now define the theory of epigenesis by means of the terminology just developed. In the theory of epigenesis there exists no *general* transformation that would carry G into M, that is, no transformation which is based on universal abstract rules and contains no reference to specific biological classes. We postulate that this is true for any description whatever. If one description existed for which this were not true, we could confine ourselves to its use and return to the theory of preformation. In terms of prediction these postulates say that there exist developmental stages, G, of the class such that knowledge of the class characteristics at this stage does not allow of adequate prediction of the class characteristics at stage M; any such prediction must involve a knowledge of class characteristics which an observer cannot find by studying the class characteristics of G.

Needless to say that within our universe of stationary biological classes there is no question of genuinely "novel information". The transformations themselves, such as the one leading from G to M, are characteristics of the class which have made their appearance already in the past history of the class. On the other hand, the study of "novel information" not represented in the history of the class would lead us straight into the theory of evolution, which is beyond our scope.

We had previously introduced the concept of a physical prediction. Now we identify such predictions with the ones obtained by integration of the differential equation of physics, possible in as much as macroscopic parameters exist and maintain their identity in such a manner that these predictions can be made in a meaningful way. On the other hand, predictions involving biotonic regularities correspond to transformations of the class description so far as such transformations are not general but involve characteristic properties of the class. We need hardly explain now that we consider the theory of epigenesis as one of the main applications of the concept of biotonic components of organismic law. It is by no means

the only one; others have been suggested earlier (Elsasser, 1958) such as the morphological stability of the adult organism, and certain functions of the central nervous system, in particular instincts and cerebral memory, but the latter will not be discussed in the present paper.

There are many biologists, especially those concerned with morphogenesis and allied subjects who are favorably inclined toward epigenetic ideas, although it is true that the principles of the theory are often expressed in an inadequate form, using such terms as "potentialities" that do not readily yield to a more rigorous definition. Again, when dealing with developmental processes one cannot overlook the tremendous body of data accumulated by biochemists and chemical geneticists which has its own intrinsic regularities. There is a good deal of evidence at present to the effect that each gene is a definite chemical structure within the DNA molecules of the chromosomes and that this structure serves as a template for the formation of a specific protein molecule, either directly or through the intermediary of a RNA molecule. These protein molecules may readily be assumed to be able to generate co-enzymes and prosthetic groups and attach the latter to themselves. In our present language we may express this by saying that there exists an isomorphism, a one-to-one mapping, between the gene system and the enzyme system of the cell. While this has not yet been proven with finality it represents an extremely likely extrapolation toward a very simple and universal rule. This means that the more general homomorphisms postulated by the theory of epigenesis do not make their appearance until we consider the relationship of the protein-enzyme system of the cell to morphological structures of the organism, be it cytological structure, or the structure of organs, or the overall organization of multicellular systems. There is no evidence, so far as we know, that would induce one to decide in favor of a simple isomorphism as between these more specifically morphological characteristics and the protein-enzyme system. As we have indicated, on the theory of epigenesis the homomorphisms required cannot be described in general terms but represent class characteristics that involve biotonic regularities.

6. Ergodization

We have pointed out that the distinction between a mechanistic theory with its purely physical predictions and a broader organismic theory cannot be made in terms of specific results of an experiment or of a limited series of experiments. Were such a decision possible, it would imply that the objects of the experiments can be considered as members of *homogeneous classes* if not throughout, then with respect to those dynamical variables specifically studied in the experiments. Instead, we have seen that the biotonic components of organismic law are based on the inhomo-

geneity and finiteness of biological classes. The dynamical relationships between macroscopic and microscopic structure which underly the biotonic aspects of organismic law can only be established in terms of inductive inferences based on observable characteristics of the class. Looking at their functional significance, we have been led to think of *design features* of organisms, which can mediate the coupling between macroscopic and microscopic variables and create and maintain the pervasive inhomogeneity of the organism. We can hardly expect that these particular characteristics of the organism just "happen", seeing that objects ordinarily encountered in inorganic nature are very far from this condition. We must rather assume that design characteristics which subserve the appearance of biotonic regularities are a significant part of the overall design of all organisms. The experimental verification of the theory presented here will have to proceed in terms of the study and interpretation of those design characteristics of organisms which maintain inhomogeneity and couple with each other dynamical variables at different levels of organization. Such components of the organism, structural, dynamical, or, more usually, both, will be designated as *ergodizers*. (The origin of the term will be explained presently.) The relationship of observational and experimental biology to the concepts developed here is then clarified: it consists in establishing and analysing the existence and function of those devices and processes in the organism that can be interpreted as ergodizers. To a considerable extent the theoretical conclusions drawn from such studies must always remain inductive since the ergodizers themselves subserve the appearance of regularities that cannot be derived deductively from the laws of physics.

There is little doubt that a great deal of evidence on this subject may be garnered from existing biological literature, but this author prefers to leave the marshalling of such evidence on the whole to biologists, in as much as he approaches what might at this time still be called the no-man's-land between physics and biology from the standpoint and with the background of the physicist. The few indications about ergodizers given are more or less random, colored by the author's lack of experience as a biologist. One thing seems clear in any event, however: metabolism in all its forms is a prime, or, rather, the prime dynamical instrument of ergodization. Like almost all structures or functions found in the organism, the functions of metabolism are not simple but composite and multiple. Evidently, some of the basic functions of metabolism are mechanistic (that is, subject to physical prediction and capable of representation by mechanistic models). Here belong for instance the processes that supply oxidizable substances and remove the end products of oxidation. But even a superficial perusal of the facts of metabolism seems to indicate that it

can only be very imperfectly explained along exclusively mechanistic lines. Thus, to quote just one example of a contradiction, the normal functioning of the brain requires a very large blood supply; but on the other hand it is well known that the energy expenditure during nervous conduction is minute. Again, the metabolic requirements of the brain during embryonic development seem to be many times larger than those of any other organ. The last-named fact may become more intelligible if we assume that the building up of the extremely complex structures of the brain is mainly an epigenetic phenomenon, meaning that these structures are only very rudimentarily preformed in the germ cells. Moreover, we may assume that the normal, adult functioning of the brain involves regularities of a strongly biotonic nature. We do not deny the possibility of a mechanistic interpretation of the facts mentioned; a mechanistic interpretation is always possible for a limited set of data, although it might radically lack plausibility in a broader context.

One of the outstanding features of metabolism *in vivo* is that chemical transformations occur commonly in the form of rather lengthy series of small steps, each step having only a very modest heat of transformation. It follows from this that these processes need not, as a rule, deviate far from equilibrium. Consider one such step, for instance a chemical reaction,

$$A + B \rightleftharpoons C + D \qquad (a)$$

with a small energy of transformation. In thermodynamical equilibrium the rate of concentrations,

$$C = c_{C+D}/c_{A+B}$$

will have a definite value, $C = C_0$. If on the other hand we start with $C = 0$, this condition may persist until we add a trace of some catalyst, whereupon equilibrium is gradually established. A living organism is not of course a thermodynamical equilibrium, it is a process. Owing to the small size of the steps involved in most of its chemical transformations it need not usually stray too far from equilibrium. Now while in a strict equilibrium the macroscopic parameters are fixed, this need not be the case in a slightly irreversible process; it is then possible for C to assume a considerable *range* of values depending on various characteristics of the system, for instance the presence of several types of catalysts with different efficiencies. These catalysts may be present in much smaller concentrations than the substances involved in process (a) and the proportions of these catalysts may in turn be determined by the chemistry of a level of organization again closer to the microscopic realm, and so on.

Return for a moment to strict thermodynamical equilibrium. Whereas its macroscopic parameters are fixed, the system is of course in constant

motion microscopically; it passes rapidly from one microscopic complexion to another, and an immense number of complexions are of course always possible. In the terminology of statistical mechanics a system in equilibrium is called ergodic if in the course of time it passes through all the complexions compatible with the given free energy of the macroscopic system. The significance of this property may be illustrated as follows: Since the energy of transformation involved in the reaction (*a*) above is small this reaction can certainly occur at equilibrium for the individual molecules, although the macroscopic concentration ratio remains constant at $C = C_0$. Now it may happen that the molecular reactions that usually take place cease to occur if all traces of catalyst are removed. Although two systems differing in their molecular reactivities may be indistinguishable macroscopically so long as they both remain in a state of equilibrium, they are of course quite different microscopically. (It is hardly necessary to say that this concept can be given a direct experimental meaning in terms, say, of radioactive tracers or isotopes. In the final process of verification the system is taken apart or must at least be removed from equilibrium.) The system in which reaction (*a*) occurs in equilibrium can assume successively an immense number of microscopic complexions not accessible to the other system. A transformation between complexions consists here in an exchange of components among some fraction of the molecules. In the language of statistical mechanics, the system where this does not occur is confined to a limited region of its phase space. The first system is ergodic with respect to the complexions engendered by reaction (*a*), the second is not.

The term ergodic has certain mathematical connotations and is rigorously defined only for a system in thermodynamical equilibrium. We can give an approximate meaning to the tendency of processes near equilibrium toward ergodicity. We may say that the system becomes ergodized if by some means (for instance by adding traces of catalysts) we can vastly increase the number of microscopic complexions accessible to it. Now in irreversible systems, especially if their structure is somewhat complicated, the *macroscopic* variables are no longer as rigidly fixed as they are for equilibria. They may vary appreciably in response to small changes of external parameters (as exemplified by conventional feedback systems). If the system is sufficiently complex there might, even for fixed external parameters, be a large number of different processes, different "paths" of the irreversible process, equally compatible with the given macroscopic parameters. How far these alternatives can be realized will mainly depend on the "ease" with which the system can move about among various groups of microscopic complexions, that is, the means of ergodization available. Now we are not usually concerned in the organism with strictly

stationary irreversible processes, but with processes that change their macroscopic characteristics slowly while not deviating very far from equilibrium. If the couplings between various levels of organization, from the macroscopic to the microscopic, are sufficiently multiple and complex physical prediction of these macroscopic characteristics may become more and more indeterminate, that is (given the limitations of measurement discussed previously) we may no longer be able to predict with any degree of confidence, on the basis of physical laws, the manner in which the macroscopic parameters of the system change. This is the necessary condition for the appearance of biotonic regularities.

We hope that this example will illustrate the concept of ergodizers. But saying that metabolic activity subserves ergodizing functions is a mere generality; such a concept will assume concrete form only if it is applied to specific cases, in biology as well as in the underlying biochemistry. We believe that in some cases at least such analysis is already within the confines of present knowledge.

Now while the dynamical characteristics of ergodization are clearly tied to metabolism, it is also possible that there exist more static structural features of organic tissue that subserve ergodization. We wish to suggest the potentialities of one such phenomenon, conspicuous by its generality. We mean the stereo-asymmetry and left-handedness of most of the molecules of biochemistry and the attendant one-handedness of the helical structures built from them. Physical experience with slightly irreversible processes (especially hydrodynamical convection and turbulence) shows clearly that such processes are capable of a vastly larger variety of dynamical configurations when the structures and forces involved are highly asymmetrical than when they are symmetrical. Symmetry tends to restrict severely the variety of dynamical processes that can be realized. Here again, we may safely assume that stereo-asymmetry, like the majority of organic forms or functions, subserves multiple and complex and not simple purposes, but ergodization might well be a major one among these. The author does not want to pursue this last hypothesis here, as he hopes to return to it on another occasion.

In concluding we wish to point briefly to two further very general aspects of the biotonic components of organismic law and of the ergodizing processes that permit their realization. Such ideas might have a bearing not only upon the more recondite aspects of biological investigations but also upon medicine. The process of healing is one in which, in our cruder language, "information" pertaining to organic structures is restored. One might well suspect that such processes cannot be understood along purely mechanistic lines but involve strong biotonic components, in which case the study of ergodizers may ultimately have considerable therapeutic

interest.† Secondly, we wish to point out the relationship between the notions of ergodization and of *homeostasis*. We have emphasized throughout the endogenous character of biotonic phenomena, and this latter they clearly share with homeostatic processes. True, much of homeostasis can be understood in terms of mechanistic models, such as relatively straightforward feedback cycles, but it is by no means certain that homeostasis can be entirely comprehended in these terms. As the feedback systems required to represent homeostatic models become more and more complex, their scheme may well pass gradually into something that is not far removed from what we have called ergodizers. The elucidation of the relationship of these two comprehensive biological concepts on an empirical basis may well lead to interesting discoveries.

Appendix

A few formulas are given here to supplement the text. Although they are not used constructively for the solution of specific problems, we hope that they might facilitate the reading of the paper for those accustomed to the language of mathematical physics.

COMPLEMENTARITY

The usual definition of the spread of a variable uses the root-mean-square average, but this presupposes a metric. A more general definition of spread applicable to any set of disjoint classes is obtained in terms of the generalized entropy. We define, assuming the classes to have equal statistical weight,

$$\epsilon = -K \sum_i p_i \log p_i \tag{1}$$

where K is an arbitrary constant. As usual, we normalize the probabilities,

$$\sum_i p_i = 1 \tag{2}$$

In the limit of a one-dimensional continuous distribution over a variable, x say, the sums are replaced by integrals. If the distribution is a Gaussian

$$p(x) = \frac{1}{(2\pi)^{\frac{1}{2}}\sigma} e^{-x^2/2\sigma^2} \tag{3}$$

we have

$$\epsilon = K \log \sigma + c \tag{4}$$

where the additive constant, c, can be made small by a suitable choice of scale.

† This was pointed out to the writer in a number of valuable conversations with Dr. R. B. Livingston of the National Institutes of Health.

We apply this to the complementarity relations of quantum mechanics for pairs of canonical variables, say position x, and momentum p,

$$\sigma_x \, \sigma_p \geq h/2\pi \tag{5}$$

where now σ_x and σ_p are general dispersions (*rms* averages) and the equality sign obtains only if both distributions are Gaussian. Using (4) with $c = 0$ we can put (5) in the form

$$\exp(\epsilon_x/K) \cdot \exp(\epsilon_p/K) \geq h/2 \tag{6}$$

Equation (6) immediately suggests a generalization of the concept of complementarity to any pair of sets of disjoint classes. Let α designate the first set of the pair and β the second set, then we say that a relation of complementarity exists between the two sets if there is a constraint,

$$f(\epsilon_\alpha) \cdot g(\epsilon_\beta) \geq C \tag{7}$$

where f and g are monotonic functions of their arguments and C is some constant.

GENERALIZED ENTROPY

This quantity is a generalization of (1), namely,

$$\epsilon = -K \sum_n \sum_i p_i^{(n)} \log p_i^{(n)} \tag{8}$$

which applies both to the thermodynamical case and to the entropy of messages (Shannon). In statistical thermodynamics, i refers to the possible occupation numbers of a cell, and n numbers the cells; in the case of a message, i refers to the possible alternative symbols for a given position in the message, and n numbers the sequence of positions. As is well known, the generalized entropy owes its superiority over other scalar measures of distribution to a number of formal properties of which the most important are that it is additive for the composition of non-correlated systems, and that it is maximized if and only if all the p_i are equal (equipartition). An excellent presentation of the applications of the generalized entropy concept to information theory has been given by Brillouin (1956).

GENERALIZED SECOND LAW

This,

$$d\epsilon/dt \geq 0 \tag{9}$$

hardly needs much explanation here. It applies to the thermodynamical entropy (Boltzmann) and to the entropy of a message which passes through a passive transducer with noise. There seem to be not as yet any significant applications of (4) or its equivalents to the case of a message operated upon by an active transducer (automaton) but the widely held view that

an automaton cannot generate "novel information" seems to indicate that it may be possible to prove significant theorems generalizing (4) to automata. In the absence of such theorems we have refrained from discussing some fairly obvious applications to the description of organism (Section 5), of ideas centering about the generalized second law.

THEOREM OF MEANS

Let the Hamiltonian of the system (expressing its gross chemical composition which we assume for simplicity as the same for all members of the class) be designated by H. As pointed out in Section 4, the microstructure of the system (class) can only be described using inductive inferences, and so the representation of the class has to be in terms of a statistical matrix, or statistical operator, ρ, which is the quantum-mechanical equivalent of the Gibbs ensemble (Ter Haar, 1954). This matrix obeys the canonical equation of motion

$$(ih/2\pi)\,\partial\rho/\partial t = H\rho - \rho H \qquad (10)$$

which stands for a set of linear differential equations that admit of straightforward integration yielding $\rho(t)$ uniquely if the initial value, $\rho(0)$, is given.

Let F be an operator representing any function of the dynamical variables of the system; its expectation value is given by

$$F(t) = \overline{F\rho(t)} = \sum_{ik} F_{ik}\,\rho_{ki}(t) \qquad (11)$$

where the last member is in matrix notation. Since $\rho(t)$ is uniquely determined by $\rho(0)$ it follows that the same is true for $F(t)$, which is the statement designated in the text as the theorem of means. Furthermore, all statistical distribution functions are determinate as functions of time; they can be subsumed under (11) either in the form of their series of moments, or, more rigorously, by means of the so-called projection operators introduced by von Neumann (1933).

Again, the initial statistical matrix, $\rho(0)$, has to be determined by means of measurements (interactions with other systems) made at times $t \leq 0$. One chooses $\rho(0)$ so that the expectation values over ρ corresponding to the measured quantities are equal to the magnitudes actually measured (Elsasser, 1937). This of course is far from determining $\rho(0)$ uniquely; for the rest it is necessary to make certain statistical assumptions (of which the equal *a priori* probability of all pure states is a typical example). Such assumptions cannot be avoided; they express indeterminacies inherent in the use of methods of inductive inference, which methods are indispensable for the description of highly complex systems, as emphasized earlier.

REFERENCES

Birkhoff, G. & MacLane, S. (1953). "A Survey of Modern Algebra". Macmillan, New York.

Bohm, D. (1951). "Quantum Theory". Prentice Hall, New York.

Bohr, N. (1934),. "Atomic Theory and the Description of Nature". Cambridge University Press, London.

Bohr, N. (1958). "Atomic Physics and Human Knowledge". John Wiley, New York.

Born, M. (1958). *Z. Phys.* **153,** 372; translated (1959) in *J. Phys. Radium* **20,** 43.

Brillouin, L. (1956). "Science and Information Theory". Academic Press, New York.

Elsasser, W. M. (1937). *Phys. Rev.* **56,** 987.

Elsasser, W. M. (1952). In "Louis de Broglie, Physicien et Penseur", pp. 87-108. Albin Michel, Paris.

Elsasser, W. M. (1958). "The Physical Foundation of Biology". Pergamon Press, London and New York.

Nordenskjöld, E. (1946). "The History of Biology". Tudor Publishing Co., New York.

Ter Haar, D. (1954). "Elements of Statistical Mechanics", Chapter 7. Rinehart, New York.

Von Neumann, J. (1933). "Mathematische Grundlagen der Quantenmechanik". Springer, Berlin; translated (1955) Princeton University Press.

J. Theoret. Biol. (1961) 1, 59–74

Methods for Measuring and Correcting the Absorption Spectrum of Scattering Suspensions

J. Amesz, L. N. M. Duysens† and D. C. Brandt

Biophysical Laboratory of the University of Leiden, Netherlands‡

(*Received* 28 *September* 1960)

It is shown that the absorbance spectrum of a scattering suspension, measured with a spectrophotometer, in which the light is collected within a small angle, is not only shifted upwards, but also flattened, as compared to the spectrum obtained with an apparatus collecting all, or nearly all, scattered light.

A quantitative description of this effect is given and, making certain assumptions, an equation is derived which makes it possible to correct for the effect of scattering. The equation is applied to correct spectra of a suspension of *Chlorella*; the results indicate that the equation describes the flattening of the spectrum of such a suspension fairly well; the best results are obtained if the light can be caught at a rather large angle.

A simple known method to collect a representative sample of the scattered light is to place a piece of light-diffusing material, opal glass or filter paper behind blank and sample vessels. In this paper a method is described for measuring such a "90° absorbance" (semi-integral absorbance) spectrum in the ultra-violet with a better precision than with a light-diffusing plate, using a fluorescing solution of esculine or sodium naphthionate behind the absorption vessels, which acts as a "perfect" diffusing layer. We have applied this fluorescence method to test the reliability of the application of several kinds of filter paper and of opal glass for measuring absorption spectra. In this way it was found that values obtained with opal glass are of rather good precision. Spectra measured with filter paper were found to be less accurate, but satisfactory for many purposes. Absorption spectra are given of *Chlorella*, *Anacystis* and *Rhodospirillum*, which were measured by means of the fluorescence method.

Introduction

Measurement of the absorbance of biological materials and its interpretation in practice presents several difficulties, because the pigments generally are inhomogeneously distributed and because these materials may considerably scatter the measuring light.

† This investigation has been aided by financial support of the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

‡Postal address: Biophysica, Nieuwsteeg 18, Leiden, The Netherlands.

In a non-scattering suspension of strongly absorbing particles, such as micro-organisms, the inhomogeneous distribution of the pigment molecules causes a flattening of such a spectrum with respect to the spectrum of a suspension of weakly absorbing particles or of a solution of the same pigment molecules (Duysens, 1956). For a scattering suspension the measured absorbance spectrum shows additional flattening, which depends to a great extent on the way of measurement. When only part of the scattered radiation is collected upon the photocell, the spectrum is flattened and shifted with respect to one obtained by collecting all, or nearly all, scattered light (cf. Latimer, 1958, Fig. 2). The effect appears to be stronger, the smaller the angle of collected light. In this paper we shall describe quantitatively the influence of scattering upon the shape of the absorbance spectrum, which has not, so far as we are aware, been done before, and we shall give a method to correct for the effect of scattering.

We shall see that, for making such a correction, it is favourable to make use of an optical arrangement which measures the light deviated at a large angle from the measuring beam. Many commercially available spectrophotometers cannot be used as such, because the photocell is too small and situated at too great a distance to collect much of the scattered radiation. They will, however, give spectra which are the same as those measured with a photocell which collects a large angle, when a "perfectly" scattering layer is placed directly behind sample and blank vessel, i.e. a layer which scatters the light rays of the measuring beam equally in all directions, independent of the angle of incidence: in that case a representative part of light, coming through the vessels, hits the photocell. Shibata and coworkers have shown that placing a piece of opal glass behind the cuvettes gives spectra that seem to be satisfactory (Shibata, Benson & Calvin, 1954; Shibata, 1958). This and other indirect evidence indicate that opal glass is a good diffusing medium. Opal glass cannot, however, be used in the ultra-violet region below about 315 mμ, because of its strong absorbancy in this region. Below 315 mμ filter paper can be used, but it is uncertain whether reliable spectra are obtained.

We shall describe a method which makes it possible to measure ultraviolet absorption spectra of scattering suspensions with a simple technique using a conventional spectrophotometer. The method is based on the fact that fluorescence is emitted equally into all directions, independent of the direction of the exciting radiation. Thus a fluorescing layer, situated behind blank and sample vessels, acts as a "perfectly" scattering plate. The method renders in principle more accurate results than methods using scattering materials, such as opal glass or filter paper, and thus can be used as a standard for testing other methods.

Theoretical

THE ABSORBANCE OF SCATTERING SUSPENSIONS

The transmittance, T_p, of a non-scattering particle is defined in the usual way as the ratio of a monochromatic unidirectional light flux transmitted by the particle, and the light flux incident upon it. The absorptance, A_p, is defined by $A_p = 1 - T_p$; and the optical density or absorbance, E_p, as $E_p = \log (1/T_p)$.

For a scattering particle analogous definitions are used. But the light flux "transmitted" by a scattering particle requires additional consideration. The light flux, I_p, incident upon the particle (Fig. 1), is partly absorbed and partly scattered in all directions. The transmittance, $T_p(\gamma)$, is

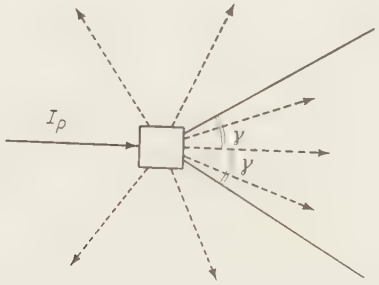


FIG. 1. Light scattering of a particle.

defined as the ratio of the light flux scattered by the particle with a deviation from the incident beam, which is less than the angle γ, and the incident light flux I_p. If $\gamma = 180°$, then $T_p(\gamma)$ is called the "integral transmittance" and $A_p(\gamma) = 1 - T_p(\gamma)$ is the "integral absorptance", that is the fraction of the incident light, which is absorbed within the particle. The integral absorptance, which can be measured by means of a light-integrating device such as a white (Ulbricht) sphere, is a quantity of importance for the calculation of photochemical quantum yields. On the other hand, this quantity is not of much use for the calculation of the intrinsic absorption spectrum of the absorbing pigment molecules within the particle, unless the light rays keep approximately their original direction. Fortunately, in most biological unicellular structures, most of the transmitted light emerges from the particle within an angle smaller than 90°, which indicates a limited deviation of the light rays within the particle from the direction of the incident beam. If this is so, then the integral transmittance is approximately equal to $T_p(\gamma = 90°)$, a quantity which, in principle, as we will see, can be determined more easily than the integral absorption.

In practice one generally measures the transmittance $T(\gamma)$, or absorbance $\log [1/T(\gamma)]$ of a suspension of many particles, instead of the values of $T_p(\gamma)$ and $E = \log(1/T_p(\gamma))$ for one particle. For non-scattering suspensions the following relation has been obtained (Duysens, 1956, equation (5)): $\ln (1/T) = p(1 - T_p)$, or

$$E = \log 1/T = p(\log e)\,(1 - T_p) \qquad (1)$$

in which e is the base of the natural logarithms and p is a constant, which is equal to the total area of all particles in the light beam, projected on a plane perpendicular to this beam, divided by the illuminated area of the vessel. For cubical particles this quantity is the same as the number of layers of particles, if the particles in the suspension are squeezed together into a solid flat layer against the front or back wall of the vessel.

The value for E given by the right side of (1) is only approximate: for cubical particles the percentage deviation from the true value is smaller than 0·6 times the percentage volume concentration of the particles in the suspension (cf. Duysens, 1956, equation (4)). For instance, for a relatively highly concentrated suspension of 3% the error in E is less than $3 \times 0{\cdot}6 = 1{\cdot}8\%$. Since p is proportional to the number of particles, E (see equation (1)) is proportional to the number of particles. Thus within the limits of validity of (1), Beer's law also holds for a suspension of particles.

To see the physical meaning of (1) let us now imagine that the pigment molecules, contained in cubical particles, are homogeneously dispersed in a solution, without changing the intrinsic absorption of these molecules. The absorbance, E_{sol}, of this solution is given by (cf. equation (1)):

$$E_{sol} = \log (1/T_p)^p = p \log (1/T_p) \qquad (2)$$

an equation easily derived, when one realizes that E_{sol} is equal to the absorbance of the p layers of particles, squeezed flat against the wall of the vessel. By eliminating p from (1) and (2), we get:

$$E(\lambda) = \frac{(\log e)\,[1 - T_p(\lambda)]}{\log [1/T_p(\lambda)]} \cdot E_{sol}(\lambda) \qquad (3)$$

In words: the absorbance spectrum $E(\lambda)$ of a suspension of particles is obtained by multiplying the absorbance spectrum of the pigments in solution by a factor, which depends only upon the transmittance of a particle. This factor is seen to approach unity, if T_p approaches unity, and to approach zero slowly, if T_p approaches zero. Thus for values of T_p close to unity, that is for weakly absorbing particles, the absorbance spectra of the suspension and the solution almost coincide, while for highly absorbing particles the absorbance spectrum of the suspension is appreciably lower than that of the solution. In other words: the spectrum of a

suspension of particles is flattened, especially at the maxima, compared with that of a solution containing the absorbing pigments (cf. Duysens, 1956, Fig. 3).

We now remark that equation (1), and also the other equations, are valid also for scattering suspensions, if we replace T_p by $T_p(\gamma)$, and E by $E(\gamma)$, provided sufficiently dilute suspensions are used. If we assume that the suspension is so dilute that most of the light transmitted by an arbitrary particle leaves the suspension without hitting another particle, then the derivation of the modified equation (4) for the scattering suspension is the same as that for the non-scattering suspensions,

$$E(\gamma) = p \text{ (log e) } [1 - T_p(\gamma)] \tag{4}$$

A deviation of the true absorbance from that given by (4) certainly occurs, if the number of particles and the amount of scattering by one particle is so great that an appreciable part of the light, which has been reflected outside the angle γ by one particle, is reflected back into this angle by other particles.

The absorbance $E(\gamma)$ of a suspension can be determined by means of a photoelectric cell or another optical device, which catches the light within the angle γ. Since equation (4) is valid for small p (small number of particles), it will remain valid for greater p in the range of concentrations, in which $E(\gamma)$ appears to be proportional to the concentration, a condition which can readily be checked experimentally. Equation (4) enables us to describe quantitatively the influence of scattering upon the shape of the absorbance spectrum. The transmittance $T_p(\gamma)$ may formally be considered to be the product of two factors:

$$T_p(\gamma) = T_p''(\gamma)\ T_p'(\gamma). \tag{5}$$

$T_p''(\gamma)$ is the fraction of the incident light that would be transmitted within the angle γ if no intrinsic absorption but the same scattering as with the real particles occurred. $E''(\gamma)$ is defined analogously. Except for that part of the scattering which is caused by the pigments themselves, $T_p''(\gamma)$ and also $E''(\gamma)$ may be experimentally determined after bleaching the absorbing pigments. $T_p'(\gamma)$ is the transmittance factor due to intrinsic absorption; it equals that fraction of the light which would be transmitted, if no loss of light by scattering occurred. Substituting (5) into (4), we obtain:

$$E(\gamma) = p \text{ (log e) } [1 - T_p''(\gamma)\ T_p'(\gamma)]. \tag{6}$$

The absorbance in the absence of scattering losses, which is generally the most important quantity we wish to determine, is equal to:

$$E'(\gamma) = p \text{ (log e) } [1 - T_p'(\gamma)]. \tag{7}$$

A usual procedure to correct for scattering is to subtract the measured spectrum of a bleached suspension from that of the original one. For

regions where no scattering due to pigments occurs $E''(\gamma)$ is equal to the absorbance of a bleached suspension; $E''(\gamma)$ is the same for an unbleached suspension and for a bleached one; it is given by:

$$E''(\gamma) = p \text{ (log e) } [1 - T_p''(\gamma)] \tag{8}$$

Correcting for scattering is often done by subtracting the spectrum of the bleached suspension from that of the original suspension. Subtracting (8) from (6) we get:

$$E^*(\gamma) = E(\gamma) - E''(\gamma) = p \text{ (log e) } T_p''(\gamma) [1 - T_p'(\gamma)] \tag{9}$$

After substituting (7) into (9) and solving for $E'(\gamma)$, we get:

$$E'(\gamma) = [E(\gamma) - E''(\gamma)] / T_p''(\gamma). \tag{10}$$

Since for scattering suspensions $T_p''(\gamma) < 1$, the spectrum $E^*(\gamma) = E(\gamma) - E''(\gamma)$ is too low; only if the angle γ is large, and the back scattering low so that $T_p''(\gamma)$ is close to unity, is a correction by subtracting justified. If $T_p''(\gamma)$ is smaller than 1, equation (10) or (11) should be used.

By substituting $T''_p(\gamma)$ from (8) in (10), we obtain the useful equation:

$$E'(\gamma) = [E(\gamma) - E''(\gamma)] / [1 - E''(\gamma)/p \text{ log e}] \tag{11}$$

Equation (11) enables us in principle to calculate the absorbance spectrum corrected for scattering, $E'(\gamma)$. $E(\gamma)$ can be measured at each wavelength; p can be calculated from the number and size of the particles. As stated, $E''(\gamma)$ is the absorbance of the bleached suspension. In a spectral region where the pigments do not have intrinsic absorption, $E''(\gamma)$ is equal to $E(\gamma)$; by extrapolation $E''(\gamma)$ may be estimated for an adjacent spectral region, where absorption does occur. In the section "Results" we have used equation (11) for the calculation of $E'(\gamma)$ spectra of suspensions of photosynthesizing organisms.

Rewriting (6), we get:

$$p \text{ log e} - E(\gamma) = p \text{ (log e) } T_p''(\gamma) \, T_p'(\gamma) \tag{6a}$$

The left-hand side is equal to the distance between the horizontal line p log e in Fig. 2 and the spectrum $E(\gamma)$. From equation (6a) it follows that: $BB'/BB'' = [T_p''(\gamma_1)/T_p''(\gamma_2)] \, [T_p'(\gamma_1)/T_p'(\gamma_2)]$. If the light rays remain approximately parallel inside the particle, $T_p'(\gamma)$ is independent of γ. If this independence of γ occurs, and if $T_p''(\gamma)$ is independent of wavelength, which may be true for a narrow wavelength region, then the quotient $BB'/BB'' = T_p''(\gamma_1)/T_p''(\gamma_2) = AA'/AA''$ is independent of wavelength, and the flattening of the $E(\gamma)$ spectra for different γ can be seen from the simple geometric construction, shown in Fig. 2.

As we shall see in the section "Results", the assumption that $T_p'(\gamma)$ is in good approximation independent of γ is valid for a suspension of the green alga *Chlorella*. The second assumption, that $T_p''(\gamma)$ is approximately

independent of wavelength in a sufficiently narrow wavelength region, is sometimes not fulfilled around an absorption maximum where $T_p''(\gamma)$, for small γ, may vary strongly with wavelength, even over small wavelength intervals (Latimer & Rabinowitch, 1959). A small-angle spectrum, then,

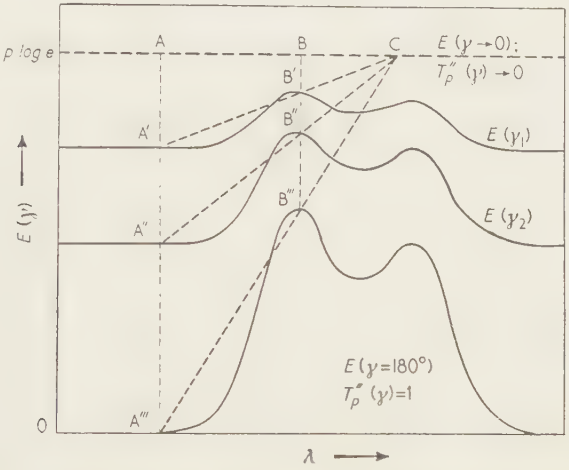


FIG. 2. Theoretical absorbance spectra of a scattering suspension, as measured at different angles. The construction of the spectra is based on assumptions explained in the text.

is considerably distorted, and it is difficult to estimate $E''(\gamma)$. However, the $E'(\gamma)$ spectrum can be calculated with better precision if a large-angle spectrum is measured, since $E''(\gamma)$ and the correction are both small.

Methods

In all our experiments a Zeiss PMQ II spectrophotometer was used. This spectrophotometer is equipped with an IP 28 photomultiplier for the region 200–600 mμ and with a vacuum photocell for longer wavelengths. The measuring beam at the place of the vessel is about 7 mm wide and 5 mm high. All vessels used were 10 mm wide (inner diameter) and 30 mm high.

Fluorescence method for measuring absorbance spectra

The fluorescence method is based on the fact that a strongly absorbing solution without self-absorption acts as a perfectly diffusing plate, because the fluorescence energy is proportional to the number of exciting quanta hitting the vessel, and the fluorescence radiation is distributed equally in all directions, whether the exciting light is diffuse or not. Vessels (Fig. 3(a)) with a fluorescing solution, placed behind both blank and sample absorption vessels, thus give, in principle, "semi-integral absorbance" (90° absorbance) spectra with very good precision.

There are, however, three practical difficulties:

(1) The fluorescing layer may transmit a few per cent of the measuring beam. The transmitted radiation would behave as if no fluorescing layer was present and give rise to a strong small-angle spectrum (which is distorted) superimposed upon the semi-integral absorbance spectrum. This radiation can be simply cut off by placing behind the fluorescence vessel a filter (Fig. 3(b)) which absorbs the measuring beam and transmits the fluorescence. For any cut-off wavelength down to about 300 mμ, such a filter is commercially available.

(2) When measuring u.v. spectra, a small amount of "false light" of longer wavelengths may be present in the measuring beam, to which the fluorescing solution may be completely transparent. This "false light" may cause a response of the photocell, which, because this light is not diffused, may be of

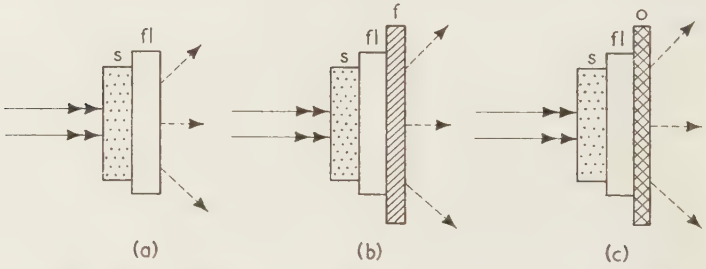


FIG. 3. Three possible methods for measuring the 90° absorbance spectrum of a light-scattering suspension: s is a vessel with suspension, fl is a fluorescing layer, f is an absorption filter and o is a piece of opal glass. For explanation see text.

the same order of magnitude as the photocell response caused by the fluorescence. When measuring the absorption at about 230 mμ by a combination of the method illustrated in Fig. 3(b) with the Zeiss PMQ II spectrophotometer we found an appreciable response, caused by false light. Filters, transmitting only a broad region in the ultra-violet below 300 mμ, to give the measuring beam a better purity are not available; nevertheless it proved to be possible to minimize the effect of false light by placing a piece of opal glass behind the fluorescence vessel (Fig. 3(c)). The opal glass diffuses the false light, so that only a small fraction of this light (for the Zeiss spectrophotometer about 1% at 340 mμ) reaches the photomultiplier. The fluorescence is weakened much less (roughly three times at 340 mμ) because the fluorescence is completely diffuse already. The opal glass diffuses or absorbs measuring light, transmitted by the fluorescence vessel, and thus has also the function of the cut-off filter of Fig. 3(b).

(3) An absorbance spectrum, measured with a conventional diffusing layer such as an opal glass, contains a contribution due to repeated back-scattering. Part of the light that passes the absorption vessel is reflected or scattered back at the opal glass, re-scattered by the suspension or vessel wall, and finally reaches the photocell. For this reason the absorbance spectrum is lowered and also, because the effect is wavelength-dependent, more or less distorted (Shibata, 1958). A correction for the effect can be made, at the expense of a

considerable loss of sensitivity, by placing a grey film between the cuvette and opal glass (Shibata, 1958). With the fluorescence method, however, the fraction of measuring light that is reflected at the wall of the fluorescence vessel is small compared with the reflected and back-radiated fluorescence light. Furthermore the wavelength distribution of fluorescence is independent of the exciting wavelength. The fluorescence reflected by the suspension is thus, for all exciting wavelengths, a constant percentage of the total fluorescence light falling on the photocell. If we determine the correction factor at one wavelength, the same correction is valid for all other wavelengths: the measured transmittance has to be multiplied by a constant correction factor smaller than 1, or a constant correction term has to be added to the measured absorbance. This correction factor can easily be determined by measuring the absorbances at one wavelength with and without a filter, which transmits the measuring and cuts off the fluorescence light between the fluorescing vessel and that containing the suspension or blank.

The method that will be described here has been worked out for use in the ultra-violet. In the region below 315 mμ opal glass cannot be used as a light diffusing plate, since it absorbs too much light. A paper diffuser can be used in the ultra-violet, but it is uncertain whether the method is very reliable; in the visible region it gives, according to Shibata *et al.* (1954), less satisfactory results than opal glass. To give a 90° absorbance spectrum, the fluorescing layer should of course be situated close behind the cuvettes, and consequently should be as thin as possible. To give the highest possible sensitivity, the fluorescence yield and the absorptance of the fluorescing solution should be close to unity in a wide region of the spectrum. The fluorescence spectrum should be located in a region of highest sensitivity of the photomultiplier (which is in the blue for an IP 28) and self-absorption should be small. Furthermore the solution should be reasonably stable.

The following fluorescing substances have been tested, dissolved in various polar and non-polar solvents, contained in a vessel of 1 mm thickness: fluorene, pyrene, chrysene, acriflavine, 5 amino-acridine hydrochloride, quinine sulphate, sodium naphthionate and esculine; the concentration was such that the absorptance was at least 0·9 in a wide spectral region.

The last two mentioned substances were found to give the best results and the following set up was finally used: Behind sample and reference cuvettes 1 mm quartz vessels were placed, containing either a 0·14% solution of esculine (Merck) or a 0·20% of sodium naphthionate (B.D.H.) in water. These solutions were stable for a few weeks at least when stored in a refrigerator. Behind the fluorescence vessels a piece of opal glass was placed. It was found that 1 cm vessels of the same concentration gave similar results, although self-absorption of fluorescence is higher.

With this fluorescence technique the spectral region above about 220 mμ can be measured; below 220 mμ the response of the apparatus is too small. In the region above 350–360 mμ the fluorescent solution gradually becomes transparent to the measuring beam but the opal glass acts as a diffuser.

Filter paper methods

The transmittance of several Whatman and Schleicher & Schull filter papers, which might be usable as a diffusing layer in the ultra-violet region, was measured with the fluorescence method. The values obtained were generally about 25% at 340 mμ, decreasing to less than 1% at 220 mμ. The Whatman filter papers tested

TABLE 1.

90° absorbance values of a Chlorella *suspension, as measured with different methods*

Wavelength, mμ	(1) Whatman No. 1		(2) Whatman No. 4		(3) Whatman No. 3 MM with paraffin		(4) Fluorescence method (Sodium naphthion.)	
	a Absorbance	*b* Sensitivity	*a* Absorbance	*b* Sensitivity	*a* Absorbance	*b* Sensitivity	*a* Absorbance	*b* Sensitivity
230	1·152	1·3	1·059	1·6	1·336	4·5	1·338	1·0
240	0·760	2·0	0·824	2·2	0·841	7·1	0·774	1·0
250	0·747	2·3	0·800	2·5	0·842	8·3	0·755	1·0
280	0·593	3·3	0·683	2·1	0·712	10·1	0·624	1·0
320	0·320	5·4	0·341	5·1	0·387	14·4	0·306	1·0
340	0·344	5·8	0·419	5·5	0·424	13·9	0·330	1·0
400	0·439	0·81	0·450	0·77	0·509	2·0	0·439	1·0

The numbers in the *b* columns give the sensitivity, compared with the fluorescence method, except at 400 mμ, where they are compared with the opal glass method. At 400 mμ the number in the last column gives the absorbance as measured with opal glass.

(Nos. 1, 3 MM and 4) had a higher transmittance at 220 mμ than the other kinds of filter paper.

We compared the fluorescence method with the filter paper method of Shibata, *et al.* (1954). Following their directions, Whatman No. 3 MM filter paper was impregnated with liquid paraffin (we used Merck No. 7149, which we found to have the lowest absorbance in the u.v. range of several samples tested) and pressed between two quartz plates to serve as a diffusing plate behind the absorption vessels. With both methods, the absorbance of a 2% *Chlorella* suspension in a 1 mm cuvette was determined. The results are presented in Table 1, together with those obtained with two kinds of non-paraffined filter paper as the diffusing layer.

Table 1 shows that oiled filter paper gives a much higher sensitivity than any other method used; it gives, however, absorbance values which are considerably higher than those obtained with the other methods. Plain filter paper gives values which do not differ so much from those obtained with the fluorescence method, except below 240 mμ where they are considerably lower. This may be caused by false light of longer wavelength, to which the filter paper is much more transparent than to the measuring beam.

The Schleicher & Schull filter papers tested (No. 589², 595, 604 and 602) gave a sensitivity which, in the region 230–250 mμ, was one to several times lower than that obtained with the fluorescence method. No. 602 gave the best absorbance values, about as good as Whatman No.1. Thus it can be seen that untreated or oiled filter paper may be used for less accurate measurements and especially when the apparatus used is not sensitive enough for the fluorescence method. All filter papers tested, however, had one drawback: we did not succeed in finding a piece of paper which was sufficiently homogeneous optically to give equal readings at the two places of the vessel carrier. The numbers of Table 1 have been corrected for this difference by running two series of measurements, one with two blank vessels, and one with a blank and a sample vessel.

Figure 4 in the next section allows a comparison of opal glass readings with those obtained with the fluorescence method. It was found that placing opal glass in our apparatus reduced the light response of the photomultiplier at 340 mμ to about 1% of its original value; adding the fluorescence vessel reduced the response about six times more.

Results

90° ABSORBANCE SPECTRA

Figure 4 gives the 90° absorbance spectrum of suspensions in water of *Chlorella vulgaris*, *Anacystis nidulans* and *Rhodospirillum rubrum*, strain 4, measured in 1 mm quartz vessels as outlined above, with an esculine solution as fluorescing layer. In the region 310–380 mμ the spectrum of the same *Chlorella* suspension is also measured with opal glass only; it can be seen that the values thus obtained are a little higher.

The correction for repeated back-scattering was found to be very small, especially in suspensions of low concentration. To determine this correction, we have placed pieces of a Schott UG 11 filter of 2 mm thickness between the fluorescence cuvettes and the sample and reference cuvettes. The filter is transparent in the region 280–360 mμ, but cuts off nearly all

the fluorescence light. The readings with the filter were compared with those without the filter, and without changing the location of opal glass and cuvettes. For a 1·25% *Chlorella* suspension of 2 mm thickness the absorbance correction thus determined was roughly 0·005 which is negligible as compared with the absorbance of the suspension which varied

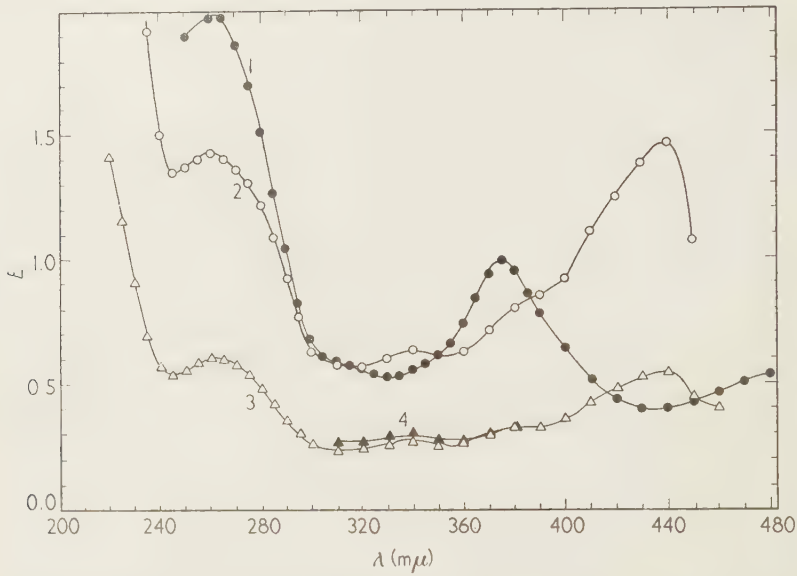


FIG. 4. Absorbance spectra, measured in a 1 mm vessel with the fluorescence method of: (1) a 5% suspension of *Rhodospirillum rubrum*, (2) a 2·4% suspension of *Anacystis nidulans*, and (3) a 2% suspension of *Chlorella vulgaris*. The absorbance of the same *Chlorella* suspension has also been measured with opal glass only (4). As explained in the text, beyond about 360 mμ the opal glass acts as a diffusing layer in the spectra presented.

between 2·0 and 0·5 in the region 230–360 mμ. For a suspension, showing less absorption in the region of the fluorescence spectrum this correction is higher.

ABSORBANCE SPECTRA MEASURED BY COLLECTING LIGHT AT DIFFERENT ANGLES

In each of two compartments of the vessel carrier of the spectrophotometer two diaphragms were placed. One with a circular hole of 3 mm diameter, which served to narrow the dimensions of the measuring beam, was placed immediately in front of the sample and reference vessel respectively. The second diaphragm, with a circular hole of 10 mm, was placed immediately in front of a sheet of opal glass and at a certain distance

behind the vessels. By varying this distance the solid angle is varied in which the scattered light is collected; e.g. by making this distance 1 cm, only light scattered with a deviation from the angle of incidence of less than 27° is collected. Figures 5(a) and 6 present absorbance spectra of a

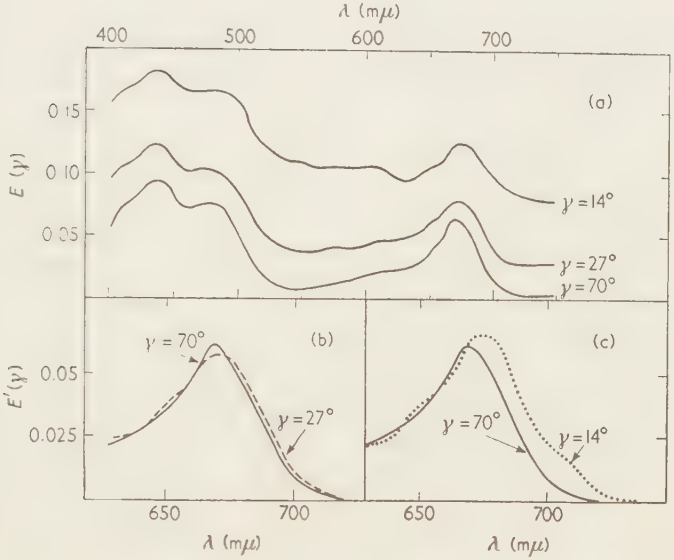


FIG. 5(a). The absorbance spectrum of a 0·30% suspension of *Chlorella vulgaris*, measured at different angles γ in a 1 mm vessel.
(b) and (c). The corrected spectra (on a different scale), in the region 630-740 mμ.

0·30% (volume) suspension in water of *Chlorella vulgaris* and of a 0·35% suspension of *Anacystis nidulans*, which are measured at different angles in 1 mm vessels.

In the region 630–740 mμ the spectra of *Chlorella* have been corrected for scattering by substituting into equation (11) the measured absorbance $E(\gamma)$ and values for p and the scattering $E''(\gamma)$ as will be explained. We have taken for p the value 0·58: if the value of p is smaller than about 0·53 or greater than 0·63, there is a clear difference between the corrected spectra. $E''(\gamma)$ has been assumed to be equal to $E(\gamma)$ at 740 mμ for the whole spectral region.

The corrected $E'(\gamma)$ spectra are shown in Figs. 5(b) and 5(c). It can be seen that for the region 630–675 mμ the corrected spectra calculated from the corresponding 27° and 14° absorbance spectra, agree well with that calculated from the 70° spectrum. In the region 675–730 mμ, the

corrected 27° and 14° spectra are too high, presumably because of strong selective scattering in this region (Latimer & Rabinowitch, 1959). The 70° spectrum appears to be only 1% flattened with respect to the corrected one. The scattering of a single particle, $1 - T_p''(\gamma = 70°)$ is only 1% since $T_p''(\gamma = 70°) \approx T_p(\gamma = 70°)$ at 740 mμ = 0·99 according to (8).

So Fig. 5 shows that after correction of the measured $E(\gamma)$ spectra, we find the same $E'(\gamma)$ spectrum for $\gamma = 70°$, $\gamma = 27°$ and $\gamma = 14°$ in the region 630–740 mμ. This indicates, according to (7), that, with the *Chlorella* suspension, $T_p'(\gamma)$ is within good approximation independent of γ, at least when $\gamma \geqslant 14°$.

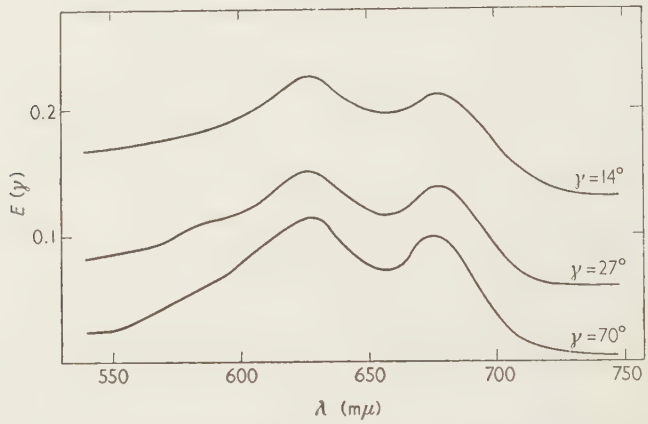


FIG. 6. The absorbance spectrum of a 0·35% suspension of *Anacystis nidulans*, measured at different angles γ in a 1 mm vessel.

From the value of p obtained the size of the *Chlorella* cells is calculated. We consider a volume of suspension contained behind 1 cm² of the front wall of the vessel. Assuming the *Chlorella* cells to be spheres with a radius r, we obtain for the total area of the particles, facing the light, $p = N\pi r^2 = (58 \pm 5) \cdot 10^{-2}$ cm², where N is the number of particles. The total volume of the cells, as determined by centrifugation in a Tromsdorff tube, was $V = 4/3\ N\pi r^3 = 3 \cdot 10^{-4}$ cm³. Dividing the two equations upon each other and solving for r, we find that the diameter $2r$ of a *Chlorella* cell is equal to $7{\cdot}8 \pm 0{\cdot}6\ \mu$, a value of the same order of magnitude as obtainable from microscopic observation.

The spectra of *Anacystis* (Fig. 6) appear to be less flattened than those of *Chlorella*. This can be explained as follows. *Anacystis* cells are much smaller than *Chlorella* cells. Since the total volume and the total scattering are in the same order of magnitude as for *Chlorella*, the scattering per cell

must be smaller for *Anacystis*, and thus $T_p''(\gamma)$ closer to unity. According to (10), an increase of deviation of $T_p''(\gamma)$ from unity means an increase of flattening.

Discussion

The absorption spectra of *Chlorella*, presented in Fig. 5, indicate that equation (11) describes the flattening of a dilute suspension fairly well, even for spectra, measured at a rather small angle. In spectral regions, however, where it is not possible to estimate the amount of scattering with sufficient precision, it is advisable to measure a large-angle absorbance spectrum, since the correction for such a spectrum generally will be small, as was the case for the 70° spectrum of *Chlorella*. The corrected spectrum obtained will then be of good precision. The corrected spectrum, $E'(\gamma)$, can be used to obtain the concentration of the pigments in the suspension only after an additional correction for flattening, due to inhomogeneous dispersion of the pigments. This correction (for cubical particles, given by (3)) has been worked out for (non-scattering) particles, in which the light rays are assumed not to deviate from their original direction (Duysens, 1956). This assumption is probably true for suspensions of *Chlorella*, and presumably for many other suspensions, as was shown by the good correspondence of the spectra of Fig. 5, which correspondence indicated that $E'(\gamma)$ was independent of γ.

An important method for studying metabolic processes in a living cell is the measurement of changes in absorbance accompanying changes in the rate of such processes. The difference spectrum obtained by plotting these changes as a function of the wavelength is flattened by intrinsic particle absorption and, especially if the light is caught at too small an angle, also by scattering in the same way as an "ordinary" absorbance spectrum. A correction for flattening is necessary, not only when conclusions are to be drawn concerning the concentration of the reacting components, but also when comparing the relative height of such a spectrum at two wavelengths, for example at the wavelengths of the α and γ bands of cytochromes.

The fluorescence method appears useful as a standard method for measuring the 90° absorbance spectrum of a light scattering sample, and thus for testing the efficiency of other methods. In principle, the fluorescing solution is a "perfect" diffuser and thus a better one than an opal plate or filter paper. The fluorescence method may be used in the u.v. region, where opal glass cannot be applied and filter paper renders less accurate results. Nevertheless, as may be seen from Table 1, reasonably good absorbance spectra can be obtained between 230 and 400 mμ with filter paper.

We have shown by comparing the opal glass method with the fluorescence method, that, as was earlier concluded by Shibata (1957; 1958) by indirect arguments, 90° absorbance spectra in the visible and near u.v region down to 320 mμ can be measured with reasonable precision by means of the opal glass method.

The fluorescence method has been used by us with a Zeiss PMQ II spectrophotometer, but it may be used with any type of sensitive commercial spectrophotometer. By means of fluorescing pigments which absorb in the visible region, it would be possible to use the fluorescence method at longer wavelengths than 350 mμ.

REFERENCES

Duysens, L. N. M. (1956). *Biochim. biophys. Acta*, **19**, 1.
Latimer, P. (1958). *Science*, **127**, 29.
Latimer, P. & Rabinowitch, E. (1959). *Arch. Biochem. Biophys.* **84**, 428.
Shibata, K. (1957). *J. opt. Soc. Amer.* **47**, 172.
Shibata, K. (1958). *J. Biochem., Tokyo*, **45**, 599.
Shibata, K., Benson, A. A. & Calvin, M. (1954). *Biochim. biophys. Acta*, **15**, 461.

J. Theoret. Biol. (1961) **1**, 75–82

The Supra- and Submolecular in Biology

Albert Szent-Györgyi

The Institute for Muscle Research at the Marine Biological Laboratory, Woods Hole, Massachusetts, U.S.A.

(*Received* 19 *September* 1960)

Current biochemistry is built on the molecular concept to which it owes most of its brilliant successes, such as the unravelling of intermediary metabolism, the isolation and chemical identification of hormones and vitamins, the crystallization of enzymes, and the establishment of the basic traits of proteins and nucleic acids, etc. In spite of these achievements, biochemistry has failed to bring us closer to the understanding of the more complex and subtle biological phenomena, like motion, nervous activity, secretion, or the establishment of concentration differences against a gradient, which all involve the transformation of chemical energy into some other form, mechanical, electric, or osmotic. These transformations are linked to the cellular structures which biochemists, interested in extracts, discard as the "residue". It is natural that biochemists should have worked first with what they could extract and bring into solution; and what is called "protein chemistry", today, should be mainly the chemistry of soluble proteins.

While the molecular concept has greatly helped to clarify the structure and function of extractable substances, it breaks down in the realm of biological structures, and its rigorous application actually retards progress.

One of the reasons that make the rigid molecular concept break down is that the electrons of one molecule may, under conditions, become located on the orbitals of another molecule, as shown by two closely related phenomena: charge transfer and conductivity. The first deals with the transfer of one electron between two isolated molecules, the latter with transfer in an assembly of a greater number of similar molecules. Both phenomena involve an overlap of orbitals which allows electrons to pass from the one to the other. Charge transfer may be symbolized by Fig. 1 in which the two sets of parallel lines on the left represent the highest filled and lowest empty orbitals of molecules D and A. If the electron clouds overlap and energy conditions are favorable, one of the two electrons of D, the donor, may pass to the empty level of A, the acceptor, the final situation being represented by Fig. 1 (b). In this state we actually have a

complex formed by two free radicals. The further reactions of the complex depend to a great extent on environmental conditions. The two molecules may remain linked together, the two electrons of D still coupled. In this case the optical spectrum of the complex may show similarities to that of the single radicals, but will give no electron spin resonance (ESR) signal. Under favorable conditions, the two molecules may part as free radicals, going into the ionic state. In this case, an ESR signal will be obtained. As

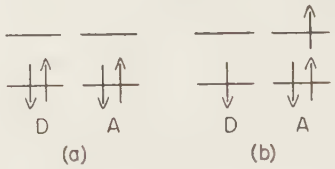


FIG. 1. Symbol of charge transfer.

has been shown earlier (Isenberg & Szent-Györgyi, 1958; Isenberg, Szent-Györgyi & Baird, 1960), biochemical pairs of biologically important substances, such as riboflavin and indoles, may form such complexes. Free radicals are most reactive but, even without going into the ionic state, the charge transfer complex D^+A^- may give unexpected reactions. The complex D^+A^- (Fig. 1 (b)), having an electron on a high lying orbital in A and an empty place on a low lying orbital in D, can be expected to be both a good electron acceptor and donor towards a third substance.

Semiconduction is schematically represented by Fig. 2 (a) and (b). In Fig. 2 (a) a number of similar molecules are brought into close proximity

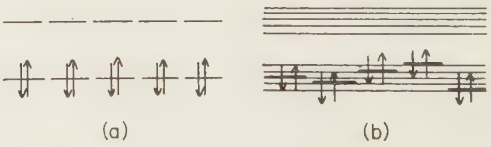


FIG. 2. Symbol of energy bands.

which allows the overlap of electronic clouds. In this situation the transition of an electron from one molecule to the other would be prevented by the Pauli principle which does not allow more than two electrons (of opposite spin) to have the same energy within the same system. But when the molecules are brought close together, they perturb one another, which changes the energy levels so that the real situation may be symbolized by Fig. 2 (b), in which the orbital energies have become slightly different. These single energy levels together form a package, a continuous band which, according to the number of participating units, may encompass a

great number of levels forming a quasi-continuous band. But, even in this case, no electrons could move in any preferred direction and the system could not conduct electricity, because all the ground levels are filled, containing two electrons of opposite spin, while the next highest energy band is empty. However, the system could be made conductant by taking away electrons from the ground level, creating "holes" in it, or by placing electrons on the empty level. Also, the system could become conductant if the empty level and filled level are close enough so that the energy of heat agitation is sufficient to raise an electron from the former to the latter, forming what is called a "semiconductor". The system would also be conductant if the single atoms forming the system had but one electron on their highest occupied level, as may be the case in metals.

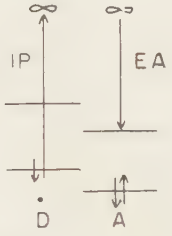


FIG. 3. Energy relations in charge transfer.

In an earlier paper I proposed that semiconduction might play a major part in energy transportation in living structures (Szent-Györgyi, 1941). The idea has not borne fruit because the lowest empty and highest filled bands are too far from one another, and the big energy quanta needed to raise an electron from the first to the latter are not available.

The way out of this situation is opened by the combination of the two reactions, charge transfer and the formation of energy bands. Let us suppose, for instance, that an electron is donated from the ground level of the system in Fig. 2 (b) to an extraneous molecule, leaving a "hole" behind. A hole having been created, the system would become electrically conductant. Similarly, if an extraneous donor would donate an electron to the empty band of Fig. 2 (b), this electron would have, here, a free mobility and could transport energy.

The energy change involved in a charge transfer reaction can conveniently be expressed by imagining the reaction taking place in two steps, in the first of which an electron is removed from D to infinity, while in the second, the electron is dropped from infinity to the empty orbital of A (Fig. 3). The energy needed to effect the first step is the ionization potential (IP) and is symbolized by the upward arrow in D of Fig. 3. The energy

gained in the second process is symbolized by the downward arrow in Fig. 3, and corresponds to the electron affinity (EA) of molecule A. The energy change would thus be $\Delta E = EA - IP + N$, N lumping up the various corrections to be made, due to interactions with the solvent, changes in coulombic forces, etc. If we use, in various charge transfer reactions different donors but always the same acceptor, then the energy change will depend mainly on the IP of the donors. The general final electron acceptor in higher biological systems is always O_2. This allows us to characterize the energy of any electron as its IP, or the "K-value" of its highest occupied orbital (Pullman & Pullman, 1952), which is a linear function of the IP. The lower the IP, or K, the less energy is needed for the transfer or the more energy will be gained by it. The lower the IP of the donor, or the higher the EA of the acceptor, the stronger the charge transfer. In a weak charge transfer only a small part of an electron is transferred, the two molecules remaining complexed and the two electrons of the original pair remaining coupled. Such complexes give no ESR signal. In a "strong" charge transfer a large part of one electric charge may be transferred, and the complex may dissociate into two free radicals which then give ESR signals.

If an aromatic structure with an extensive conjugated system interacts with another molecule of similar structure, the π pool of the first may donate an electron to the π pool of the latter in a so-called π–π interaction. This will be, for instance, the case in quinhydrone, where the π system of hydroquinone may transfer charge to the π system of the quinone.

Not only can conjugated systems act as donors. Atoms like O or N, having a "lone pair" of electrons, can donate one of their non-bonded electrons, acting as "local" electron donors.

The study of indole has led to the recognition of still another possibility. The indole derivatives interest the biologist because various biologically active substances, as growth hormones (indoleacetic acid), regulators (serotonin), or amino acids (tryptophan) belong to this group. It has been shown earlier that serotonin can form a strong charge transfer reaction with riboflavin. The result is contrary to expectation, since, with this relatively high K-value, indole cannot be expected to be a strong electron donor. If indole is made to react with one of the classical electron acceptors, like sym-trinitrobenzene (TNB), it actually behaves as a rather poor donor. However, when iodine is used as an acceptor, indole gives with it a black precipitate, which gives a very strong and narrow spin resonance signal, the g value of which is close to that of a free electron, indicating that a strong charge transfer has taken place (Szent-Györgyi, Isenberg & Baird, 1960). The I_2 molecule is not an especially strong acceptor. It is of about the same order as trinitrobenzene. The difference between the two

is that the I_2 molecule is relatively small, so it can also act as a "local" acceptor, interacting also with single C atoms or pairs of neighbouring C atoms. The calculations of the Pullmans (1952) show that the carbon in position 3 has an especially high electron density, while its neighbour, C2, has but a somewhat lesser charge. These two atoms, conjugated with the whole π pool, are thus capable of acting as "local" donors, while drawing from the π pool for the electron to be donated.

As has been shown by the Pullmans (1955), carcinogenic hydrocarbons also have in their so-called "K-region" a pair of C atoms of high electron density, a pair which makes part of an extensive conjugated system.† It was found that the strongly carcinogenic hydrocarbons gave similar strong black charge transfer complexes with I_2, and the question occurred whether carcinogenesis is not connected with the ready donation of an electron in a "local" reaction. If so, then also carcinogens, which belong to other groups of substances and have no well-defined "K-region", should also give strong charge transfer complexes with I_2 which can be recognized by their black color. This was actually found to be the case. Hitherto, all strong carcinogens tested have given such a reaction (fluorenes, diphenyls, naphthylamines, azobenzenes), while their closely related non-carcinogenic homologues, or isomers, gave no such reaction.

The question arises—why, then, is indole not carcinogenic? All the carcinogens tested gave a strong charge transfer with TNB, as indicated by the dark red or purple color of their charge transfer complex, while indole gave with TNB only a weak charge transfer, as indicated by the faint yellow color of its charge transfer complex.‡ TNB indicates the tendency of the whole π pool to part with one of its electrons, as expressed, also, by the K-value of that molecule. Summing up, this experience suggests that a substance becomes carcinogenic if it is capable of giving off an electron in a "local" charge transfer, which is backed up by a strong donor tendency of a π pool. Why high polycyclic hydrocarbons like naphthacene, perylene, or violanthrene, which have a low K value and give strong charge transfer with I_2, are not carcinogenic is a different question. This may be due to their poor solubility or to strongly developed "L" regions, which, according to the Pullmans (1955) antagonizes carcinogenicity.

† The "K-region" should not be confused with the "K-values" discussed earlier in this paper. The K-values indicate the energy of the highest filled molecular orbital, while the K-region denotes two specific C atoms in aromatic hydrocarbons.

‡ The yellow color indicates that blue light was absorbed, that is, that the relatively high energy of blue light was needed to transfer an electron. A red or purple color means that light of relatively long wavelength is capable of transferring electrons and the curves of Fujimori (see Szent-Györgyi, 1960, p. 63) indicates that in this region spontaneous transfer becomes possible, that is, no light quanta are needed to transfer electrons.

This study of I_2 complexes shows that abstruse-looking quantum mechanical considerations may lead to interesting experiments bearing on important and urgent problems of current biochemistry. To spin this yarn farther, we might ask what happens to an electron donated to an empty energy band (Fig. 2 (b)). Such an empty band can, in all probability, be found in proteins with their H-bridges (Evans & Gergely, 1949; Cardew & Eley, 1959). In such a system the electron moves freely, belonging thus equally to the whole system and to all the molecules taking part in the building of that system. We do not know how extensive these systems are within the cells. Maybe, within the structural proteins, they extend over the whole cell, making one single electronic system of it, and so giving a deeper meaning to the cellular concept. The electron passing from a donor molecule to an acceptor molecule (Fig. 1 (b)) may also move on from the latter to the empty level of a third, from there to a fourth molecule, cascading down gradually to the level of O_2, the final biological electron acceptor, yielding gradually its energy, which then could be used by the cell for its function or maintenance. Such a fall, involving but one electron, does not entail any covalent changes. The underlying structure only forms, so to say, a quantum mechanical framework, in which these electronic changes can take place. So we arrive at biochemistry without chemistry if, by chemistry, we mean a rearrangement in molecular structures.

Possibly this subdivision of biochemistry into a chemistry of structures and soluble cellular components reflects a deeper regularity, however crude this division may be. As is generally known, the final source of all energy, driving life, is the radiation of the sun. The photon, interacting with matter on this globe, raises an electron to a higher energy level. As a rule the electron drops back to the ground level within a very short period, of the order of 10^{-8} secs. Life has learned to catch the electron in its excited state and utilize its excess energy. It seems likely that it was this excess energy of single electrons which in the beginning drove life and is still driving it today. The later developments chiefly concerned the storage and transport of this energy in the form of bond energy. Grossly speaking, the soluble cell constituents seem to be concerned, mainly, with this stabilization and *transportation* of chemical energy, while the structures which eventually use this energy, are concerned with the *transformation* of this bond-energy into the different sorts of work mentioned at the outset. My research work is led, at present, by the supposition that the bond-energy is reconverted eventually into the energy of single electrons when it has to interact with structure, maintain it in the "living state" (Szent-Györgyi, 1960), or produce those various sorts of work which underly the subtle and complex biological phenomena, such as motion, consciousness, thinking, etc. The function of soluble proteins can be duplicated *in vitro*

and expressed with symbols of classical chemical chemistry, letters and dashes, involving covalent changes, that is, changes in electron pairs, while the working of structures involves changes in distribution and energy levels of single electrons.†

Naturally, the building of protein structures involves covalent chemistry, involves the bonding of single molecules as the building of a house involves placing bricks side by side and linking them with mortar. But, once the house is finished, the bricks, as units, disappear and new "supramolecular" units, as walls, emerge, while the function of the wall is dominated by the "submolecular" qualities as thermic and mechanical properties of silicate particles of which bricks and mortar are made.

The duality of soluble and structure proteins may reflect not only the story of life, but also the story of biochemistry. When biochemistry went into bloom at the end of the last century the atom was an indivisible unit, molecules the aggregates of such units. Biochemistry was thus based on the rigid molecular concept. Later it was recognized that the atom is a whole universe and chemical reactions may be but the overall result of a series of subtler changes within these systems (e.g. polarization, intramolecular shift of electrons, etc.). This is broadening our outlook at present, leading to a better understanding of the mechanism of molecular reactions, as those of enzymes or coenzymes. The subtler and more complex biological phenomena, linked to structures, may necessitate a further extension of our outlook, a fusion of the molecular with the sub- and supra-molecular.

REFERENCES

Cardew, M. H. & Eley, D. D. (1959). *Disc. Faraday Soc.* No. 27, 115.

Evans, M. G. & Gergely, J. (1949). *Biochim. biophys. Acta.* **3**, 188.

Isenberg, I. & Szent-Györgyi, A. (1958). *Proc. nat. Acad. Sci., Wash.* **44**, 857.

Isenberg, I., Szent-Györgyi, A. & Baird, S. L. Jr. (1960). *Proc. nat. Acad. Sci., Wash.* **46**, 1307.

Pullman, A. & Pullman, B. (1955). "Cancérisation par les Substances chimiques et Structures moleculaires," Masson, Paris.

† One may wonder how single electrons, with changes in their location, could induce those extensive changes in the physical state of living systems which go with function. This question can, at present, be approached only by speculation. However, it is not impossible that the principles of muscle activity represent a more general principle of biological function. We know from immune reactions, that the interaction of two specific proteins may induce profound changes in physical state (e.g. precipitin reactions). It has been shown in my earlier laboratory that muscle contraction comes about by the interaction of two proteins, actin and myosin. It has also been made probable that these two proteins, in resting muscle, are kept apart by repulsive forces. The repulsive forces are, in all probability, coulombic forces, decreasing with the second power of the distance, while the attractive forces are, to a great extent, van der Waals forces which vary with a much higher power of the distance. This must lead to a very subtle balance which can easily be disturbed, at least at one point by an electron, starting up a "zipper process".

Pullman, B. & Pullman, A. (1952). "Les Théories électroniques de la Chimie organique," Masson, Paris.
Szent-Györgyi, A. (1941). *Science* **93**, 609.
Szent-Györgyi, A. (1960). "Submolecular Biology," Academic Press, New York.
Szent-Györgyi, A., Isenberg, I. & Baird, S. L. Jr. (1960). *Proc. nat. Acad. Sci., Wash.* **46**, 1444.

J. Theoret. Biol. (1961) **1**, 83–97

An Analysis of the Idea of "Resources" in Animal Ecology

H. G. Andrewartha and T. O. Browning

Unit of Animal Ecology, Department of Zoology, University of Adelaide, Australia

(*Received* 30 *September* 1960)

1. Introduction

Elton (1927) told how the small island of Berlenga supported a population of rabbits until some cats were brought to the island. At first the cats thrived, but in due course they ate all the rabbits and then the cats died from starvation. The island of Berlenga is small and uniform and surrounded by an impassable barrier. Elsewhere in the world cats and rabbits live side by side, but in places that are larger, more heterogeneous, and from which escape is possible. A number of laboratory experiments have been done in which populations of predator and prey have been brought together in a small uniform place. The usual result is that one or both species have died out (Gause, 1934; Gause, Smaragdova & Witt, 1936; Utida, 1957). In nature the same result may be observed in local situations that need not be islands. But in general most species are distributed over such large and heterogeneous areas that an ecologist is unlikely to come across the spectacle of a whole species, or indeed any extensive population, dying out because it has eaten all its food. Nevertheless, in many of the natural populations that have been studied, shortage of food, living-room, nesting-sites or some other material necessity of life has been shown to be an important hazard reducing the animal's chance to survive and multiply. It is convenient to have a single name for all the material necessities of life; in this paper we shall call them "resources".

Shortages that arise from overcrowding in local situations are obvious and have often been described. Along the seashore in almost any part of the world rocks, piles and other such objects that are uncovered at low tide are often seen smothered with animals, providing clear evidence of an extreme shortage of living-room (McDougall, 1943). Outbreaks of insect pests such as locusts and armyworms provide striking examples of animals whose numbers have become great relative to their stocks of food in local situations. Among the vertebrates it is perhaps the ungulates that are most

likely to increase in numbers to the point where they eat themselves out of food (Errington, 1946).

Jackson (1937) was probably the first to draw attention to the circumstances in which most of the individuals in a population may go seriously short of food even though they are surrounded by more than enough food to sustain the whole population. This is likely to happen if the food is so *inaccessible* that an animal has little chance of finding enough. Potts & Jackson (1953) brilliantly demonstrated this principle at Shinyanga when they exterminated two species of tsetse flies from 600 square miles by shooting most, but not all, of the "game" in the area. Andrewartha & Birch (1954) discussed this principle under the heading "relative shortage of food" but they did not give it the prominence it now seems to deserve.

They called the other sort of shortage which is immediately concerned with excessive numbers of animals relative to their stocks of food an "absolute shortage of food". It seems profitable to examine these ideas more closely.

2. Absolute Shortage

Sheep grazing a lush pasture, or a few caterpillars feeding on the leaves of a large tree are surrounded by plenty of food that is entirely accessible to all of them—there would seem to be little chance that any would fail to get all the food that it needed. The farmer's careful husbandry may prevent the sheep from overgrazing their pasture; and one might expect some other component of environment to do the same by the caterpillars. (It is the common lot of most species to remain rare relative to their food (Andrewartha & Birch, 1954, p. 22).) But in the event that these restraints are lifted the animals might be expected to multiply until they are too numerous for their supply of food. Then some may die without reproducing, not because food has become inaccessible but simply because there is not enough to go round.

There are three statements that may be made about this situation, two of which seem to us to be relevant to ecology. We may say that the animals experience an *absolute shortage* of food (or other resource). This statement emphasizes the fact that some or all of the animals fail to get enough of the resource because there is not enough to go round. It is profitable to look at the matter in this way when we are chiefly interested in the animals in the present generation, i.e. in their chance to survive and produce offspring. Alternatively we may say that the population of animals has become *so numerous that a large proportion of the resource is being consumed.* This statement focuses attention on the future of the resource. For example, if the resource is a growing plant and many animals are eating it, there may be little chance of the plant regenerating to provide food for

future generations. Finally we may say of this situation that *intraspecific competition* is taking place. This statement emphasizes the fact that some animals are better than others at getting enough of a resource that is in short supply and so have a better chance of contributing progeny to the next generation. This idea, being chiefly relevant to the study of the evolution of the population, enters only indirectly into the scope of ecology.

But the ideas of "density relative to resources" and of "absolute shortage" of resource are relevant to certain problems of population ecology. For example, the four case-histories which we have outlined below show how different sorts of resource may vary in their response to use by a dense population and how animals may vary in their response to an absolute shortage of resource.

(i) The blowfly *Lucilia sericata* may be reared in the laboratory on meat. If the number of maggots on a given quantity of meat is carefully regulated (so that there are enough to eat all the meat, yet not so many that there is too little meat to go round) it is possible to rear about 1,600 well grown maggots from one kg of meat. However if the density of the population is increased a little there will be an absolute shortage of food, yet some maggots may still be able to complete their life-cycle. If the density is increased still further until the absolute shortage becomes extreme, all the food may be consumed without any maggot getting enough to complete its life-cycle (Ullyett, 1950). In these circumstances all the food is "wasted" because none is used to produce parents for another generation.

In nature the number of maggots that are found in a carcase often exceeds the optimum. Waterhouse (1947) exposed the carcases of 27 sheep near Canberra and allowed the wild flies to lay eggs on them freely. Many more eggs were laid than the carcases could support and the absolute shortage of food was extreme. The carcases produced from 109 to 67,000 adult blowflies with an average of 10,000. At the optimum rate mentioned above a carcase weighing 50 kg might have produced 80,000 adult flies. The difference may be attributed to the severe absolute shortage of food. A number of species were using the carcase at the same time and this complicates the full explanation of the results but there can be no doubt of the severe absolute shortage of "effective" food (see footnote to Table 1) for that generation of maggots, and its influence on the size of the next generation.

The supply of food for the next generation depends on the number and size of carcases in the area. This is determined by the number of deaths among sheep and other suitable animals in the area which is, of course, quite independent of the density of the population of maggots in the carcases that bred the last generation. In this respect blowflies are typical

of a large number of species that eat carrion, refuse, or "litter" of any sort, or that use such material for building nests or for some other purpose. For such species the supply of resource for future generations is not influenced by the density of the population of animals using the resource in the present generation; but the size of the next generation depends upon the absolute shortage experienced by the present generation (see section B (*a*) of Table 1).

(ii) Leopold (1943) found that the numbers of deer on the Kaibab plateau in Arizona were determined largely by predators until most of the predators were shot. At first the deer increased greatly in numbers and almost ate themselves out of food. At the same time much of the herbage that had been grazed and trampled did not regenerate. The deer became fewer but did not die out because there was always *some* food to be found. Their numbers came to be determined largely by the distribution and abundance of their food and especially its accessibility. In this case the numbers of animals in the present generation, by virtue of their feeding, determined the distribution and abundance of the food for future generations; and the absolute shortage of food experienced by the present generation reduced the number of young born into the next generation. That is, the deer responded to an absolute shortage of food in the same way as the blowflies; but their food responded to the number of animals using it unlike the carrion which was food for the blowflies (see section A (*a*) of Table 1).

(iii) The bug *Nezara viridula* used to be a serious pest of beans, tomatoes, and other crops in parts of South Australia until the egg-parasite *Microphanurus basalis* was introduced and became established. The females of *Microphanurus* rarely oviposit into an egg of *Nezara* that already contains an egg or larva of *Microphanurus*. Even if they do, only one larva will survive because the other gets eaten along with the contents of the egg. This behaviour ensures that no matter how great the absolute shortage of eggs of *Nezara* may be relative to the numbers of *Microphanurus*, nevertheless none will be wasted on individuals that do not get enough food to complete their life-cycle. So the numbers of individuals of *Microphanurus* in the next generation does not depend on the absolute shortage experienced by their parents but merely on the amount of food that was present. In this respect *Microphanurus* differs from both blowflies and deer. The food of *Microphanurus* resembles that of the deer (but differs from that of the blowflies) in that the supply of food for future generations depends on the numbers of *Microphanurus* seeking food in the present generation.

(iv) According to Kluijver (1951) a population of tits *Parus major* living in a mixed woodland of young trees were few relative to another

population living nearby in a woodland of old trees. The young trees had few holes suitable for nests. Nearly all the holes were used so, despite their low numbers, the tits must be considered to be experiencing an absolute shortage of nesting-sites. The territorial behaviour of the birds ensured that a pair that managed to obtain a nest had the unrestricted use of it during the breeding season. Thus each nesting-site that was used at all was used to rear potential parents for the next generation. In this way the tits' response to an absolute shortage of a resource resembles that of *Microphanurus* and differs from that of both deer and blowflies. Because the nesting-sites are used without being destroyed the supply of nesting-sites for the next generation is independent of the number of tits that sought to use them in this generation. Nesting-sites for tits thus resemble food for blowflies and differ from food for both deer and *Microphanurus*.

The relationship between the codlin moth and its food is similar to that between *Parus* and its nesting-sites because (*a*) its presence in an apple does not influence the size of the crop in the following year and (*b*) it behaves in such a way that any apple that has one or more eggs laid on it is likely to give rise to one, but not more than one, mature caterpillar (Andrewartha, 1957). Certain sorts of parasites also behave in this way.

The case-histories that we have described above may be classified as in Table 1. We think that this classification may apply quite generally to all natural populations that become abundant relative to any sort of material resource.

TABLE 1.

A. The supply of resource for future generations depends on the number of animals in the present generation using the resource.	
(*a*) The amount of resource used effectively† by the present generation depends upon the absolute shortage that they experience.	(*b*) The amount of resource used effectively by the present generation is independent of the absolute shortage that they experience.
Deer on Kaibab, *Cactoblastis*, cats on Berlenga.	*Microphanurus*.

B. The supply of resource for future generations is independent of the number of animals in the present generation using the resource.	
(*a*) The amount of resource used effectively by the present generation depends upon the absolute shortage that they experience.	(*b*) The amount of resource used effectively by the present generation is independent of the absolute shortage that they experience.
Lucilia, certain parasites, most "detritus feeders".	*Parus*, codlin moth, certain parasites.

† "Effective" is used in the sense defined by Andrewartha & Birch (1954, p. 498), i.e. it refers to that part of the resource that is actually used by those individuals that get enough to mature and so have a chance to contribute to the next generation.

3. Relative Shortage

Species like *Parus* and *Lucilia* which belong to category B of Table 1 do not, by virtue of their nesting (*Parus*) or feeding (*Lucilia*), influence the distribution and abundance of these resources for the next and subsequent generations. But species of category A do; *Cactoblastis cactorum* is such a species.

When *Cactoblastis cactorum* was first let go in Queensland in 1925 the insects were surrounded by a great abundance of food—more than thirty million acres densely covered with *Opuntia* spp., carrying about 500 tons of *Opuntia* per acre. Dodd (1940) described how the great abundance of food for *Cactoblastis* was changed, as a result of their feeding, to a widespread absolute shortage; over extensive areas there were too many caterpillars for the amount of food present. Many died from lack of food but, unlike the cats and rabbits on Berlenga, neither *Cactoblastis* nor *Opuntia* was exterminated over their entire range. The climax was described briefly as follows by Nicholson (1947): "The end result, which still persists, is that prickly pear is scattered in small isolated groups, with wide intervals between them. In a few of these *Cactoblastis* is able to increase rapidly in numbers on them. In the other groups of prickly pear which have not so far been found by *Cactoblastis*, the pear tends to spread, but sooner or later is found by the moth, and the destruction of these groups is achieved shortly afterwards. In the meantime seed is scattered in new places, so maintaining the existence of prickly pear." The important point in this situation for our purposes is the change that occurred in the distribution of the food as a result of the feeding of *Cactoblastis*. From being abundant and widespread *Opuntia* became relatively rare and patchily distributed.

Similarly when the coccinellid *Rodolia cardinalis* was introduced into the Californian citrus groves which were threatened with destruction, so numerous were the coccids *Icerya purchasi* feeding on the trees, it increased greatly in numbers at first, destroying most of the *Icerya*. As a result, the distribution of *Icerya* changed in much the same way as that of *Opuntia*. From being widespread it came to be extremely patchily distributed, in small colonies which flourished for a time until they were found by *Rodolia* and destroyed. The population of *Rodolia* fell to quite low numbers and has remained so, being controlled mainly by the distribution of its food (Smith, 1939; and see Andrewartha & Birch, 1954, p. 474). A similar situation resulted from the feeding of the deer on the Kaibab plateau.

Cockerell (1934; also quoted in Andrewartha & Birch, 1954, p. 23) described another situation in which small colonies of scale insects feeding on mesquite and other bushes in New Mexico play "hide and seek" with their natural enemies. When a colony is found it is usually exterminated,

but young larvae, blown by the wind or carried on the feet of birds, escape and establish new colonies which flourish until they too are discovered and destroyed. The result is very like the other three cases we have considered but, whereas in them the whole process was observed and described, in this case we have no information on how the situation came about.

In the cases of *Cactoblastis*, *Rodolia*, and the deer the progress of events was similar. At first food was plentiful and the animals multiplied exceedingly, until an acute absolute shortage of food occurred and the population declined to quite low levels. But because the area of distribution of the food was large relative to the powers of dispersal of the feeding animals, the nature of the distribution of the food changed, becoming patchy where once it had been virtually continuous. It thus came to resemble the distribution of coccids described above. Under these circumstances much food goes undiscovered for a time even though many animals do not get enough to eat. This state of affairs has been called a "relative shortage of food" by Andrewartha & Birch (1954). The history of any *local* colony of feeding animals resembles that of the cats on Berlenga island if we abstract from them the possibility of migration but it is the relative shortage of food of the population as a whole that seems to us to be the essential feature of the situation.

The powers of dispersal of the animal relative to those of its food seem largely to determine the density that the animals will maintain. When the powers of dispersal of the animal are small relative to those of its food much food will go undiscovered and the food will be abundant, whereas the animal may be abundant locally but will be very patchily distributed. But when the powers of dispersal of the animal are great relative to those of its food then both the animal and its food are likely to be rare and very patchily distributed and the animal is likely to be regarded either as a serious pest or as a valuable agent of biological control, depending on the point of view of the observer.

The observations of De Bach, Fisher & Landi (1955) and De Bach & Sisojevic (1960) on the scale-insect *Aonidiella aurantii* and its predators *Aphytis chrysomphali* and *A. lingnanensis* illustrate these principles. The population of *Aonidiella* that lives on citrus near the coast of California and is preyed on chiefly by *A. chrysomphali* remains scarce and patchily distributed in much the same way as Cockerell described for the coccids on mesquite. But inland in southern California where *Aonidiella* also lives on citrus and is also preyed on by *Aphytis*, but chiefly *A. lingnanensis*, the winter is colder than near the coast; and cold winters kill many of the predators without harming many of the scale insects. For a period after a severe winter the scale insects become abundant and there is little shortage of food for *Aphytis*. In the most inland area that De Bach *et al.* studied

severe winters were frequent and the few *Aphytis* were usually surrounded by an abundance of food. In the coastal area it was not *Aphytis* but *Aonidiella* that was surrounded by an abundance of food. Inland the predators were kept rare relative to their food by the occurrence of severe winters; near the coast the prey were kept rare relative to their food by predators.

In an area which is intermediate between the inland and the coastal areas that we have so far discussed severe winters recur less frequently than inland. During intervals between severe winters *Aphytis* increased, sometimes to the point where it destroyed most of its food. But, as we have come to expect in these circumstances, a few *Aonidiella* escaped discovery while many *Aphytis* died without finding food for their progeny. It is characteristic of this sort of relative shortage, and indeed of all those that we have discussed so far, that the relative shortage arises as the food becomes scarce, the scarcity of the food being caused by the animals feeding abundantly upon it.

Relative shortages of food (or other resource) that do not depend on this interaction between the animals and their food (or other resource) and are therefore independent of the density of the population may also be found in nature; and this sort of relative shortage may happen to any species whether it belongs to category A or B of Table 1. Such relative shortages, being largely dependent on other components of environment, may arise in diverse ways; but some of this diversity depends on the differences between species that are contrasted in categories (*a*) and (*b*) of the Table. We shall mention several examples that illustrate this diversity without exhausting it.

The food of certain mosquito larvae consists of minute particles that are wafted into the mouth on a current of water that is maintained by the vibration of "feeding brushes". If the concentration of particles in the water falls below a certain threshold the larvae may die from starvation. This would happen if there were only one larva in a large pool even though it contained enough food, were it concentrated, to support many larvae. The shortage in these circumstances is of the sort that we have called relative, yet it may clearly be independent of the density of the population.

According to Main, Shield, & Waring (1959) quokkas *Setonix brachyurus* (small marsupials) which live on the island of Rottnest off the coast of Western Australia, become debilitated in summer and many die, especially during periods when the rainfall is below average. The death-rate is greatest among immature animals. A variety of physiological measurements point to the conclusion that the chief cause of death is shortage of food, especially shortage of protein. Yet the area where the quokkas live is covered by an impenetrable thicket of various species of shrubs which the quokkas eat. During part of the wet season (which in this climate occurs

during winter) the shrubs put forward young growth which is probably rich enough in protein and water to provide most of the quokkas' needs for growth and reproduction. But during the rest of the year, especially during the dry, hot summer, the plants remain dormant and sclerophyllous, and the quokka is limited by time and the size of its stomach from eating enough of this malnutritious diet to get enough of the nutrients it needs. In part of the island their condition is aggravated by the absence of drinking-water and they have to obtain most of their water by eating succulents, notably *Carpobrotus*, which are even poorer in protein than the sclerophyllous shrubs that form the rest of their diet. Main *et al.* were unable to discover any other component of environment that seemed important and concluded that the numbers of the quokka were determined largely by shortage of food. The relative shortage of food experienced by the quokka is largely independent of the density of the population because the nutritious food becomes malnutritious with the advance of the seasons.

In this last respect the relative shortage of food experienced by the quokka resembles that described by Andrewartha (1957) for the grasshopper *Austroicetes cruciata*. During the spring the grasshoppers eat annual grasses that flourish during winter and spring but wither and die as the hot dry summer develops. As the season advances small tufts of grass or small parts of the tufts remain green after most of the rest has withered. But these edible pieces are so sparsely distributed that the last grasshopper dies from starvation long before the last edible piece of grass has been eaten.

The tsetse fly *Glossina morsitans* sucks blood chiefly from the larger species of antelopes. It requires to feed frequently, especially during hot weather but it does not stay near its food after engorging. So each meal is preceded by an independent search. If antelopes are sparsely distributed, a fly may have little chance of finding enough food to produce many or any offspring before it dies, despite the fact that the amount of blood in even one beast would be enough for many flies. Jackson (1937) pointed out that there would be a particular density of antelopes that would enable the flies to maintain births just equal to deaths. Changes in the number of antelopes would occur quite independently of the number of flies feeding on them but would nevertheless control the rate of increase or decrease in the population of flies. *Glossina* is like *Aphytis* to the extent that each meal is preceded by a search in which there is a finite chance of not finding food, but unlike *Aphytis* it does not influence the abundance of food for the next generation.

The sheep tick *Ixodes ricinus* is like *Glossina* in that it engorges with blood each time it finds a host. It needs three such meals in order to complete its life-cycle. When "searching" for a meal the tick waits on a grass

stem for a sheep to brush it in passing. Its chance of finding a meal depends on the density of the population of sheep but is independent of the density of its own population. Milne (1947) estimated that on one farm that carried about one sheep to the acre the odds were about 20 to 1 against a tick finding the three meals it needed to complete its life-cycle.

According to De Bach (1949) the mealybug *Pseudococcus adonidum*, living on citrus in California, is preyed upon by an "internal parasite" *Anarhopus sydneyensis*. The females of *Anarhopus* pierce the integument of the mealybug and lay an egg which develops into a maggot inside the body of the mealybug. There is also present in the area a "hyper-parasite" *Lygocerus* sp. which uses as food, during its larval stage, the larva of *Anarhopus*. The female *Lygocerus* probes the body of the mealybug with her ovipositor; if she finds a larva of *Anarhopus* she lays an egg into it, if not she withdraws her ovipositor and tries elsewhere. She requires about 15 minutes to "explore" one mealybug and seems to expend considerable energy in the course of her search; this limits the number of mealybugs that she may explore in her lifetime. At certain seasons of the year the mealybugs may be abundant but only a small proportion of them may contain a larval *Anarhopus*. Under these circumstances many *Lygocerus* may fail to find a larval *Anarhopus* and so fail to contribute progeny to the next generation even though *Anarhopus* may be more abundant than *Lygocerus* and many of them go undiscovered.

The case-histories that have been given above show that the idea of relative shortage of a resource is more complex than the idea originally analysed by Andrewartha & Birch (1954). We might arrange the six species in something of a graded series which shows that the behaviour patterns which served to classify species into the (*a*) or (*b*) categories of Table 1 (when we were discussing absolute shortage) are also important in discussing the animals' responses to a relative shortage. The mosquito larva (and other "filter-feeders") stand at one extreme. They experience a relative shortage because the food becomes too dilute—they might still not get enough even though they were able to eat all the time. The quokka is rather similar but it is limited in the time that it can spend eating. The filter-feeders and the quokka are like the deer and the blowflies of Table 1 in that most individuals will probably get some food, though perhaps not enough to grow up, no matter how severe the shortage may be. At the other extreme *Lygocerus* experiences a relative shortage because its food is hidden from it. But if *Lygocerus* finds any food there is sure to be enough to rear one individual for the next generation. In this *Lygocerus* is like *Parus* and *Microphanurus* in Table 1. The tsetse fly and the sheep tick are intermediate between those extremes. Both can count on getting a full meal if they get any at all; but the tick requires three meals to produce any

offspring, whereas the fly requires many meals to produce one or a few offspring. Species that come near the extreme represented by the mosquito larva would seem to be more vulnerable to increasing degrees of relative shortage than are those that come near the other extreme.

There is another complexity in the idea of relative shortage which is brought out by these case-histories but which was not apparent in the original statement. The essential difference between *Cactoblastis* and *Glossina* is that the scarcity of prickly pear is caused by *Cactoblastis* feeding abundantly on it whereas the scarcity of antelopes arises from causes which are quite independent of the feeding of *Glossina*. Furthermore, local populations of *Cactoblastis* frequently consume all the food that is immediately available to them and experience an acute absolute shortage as a result; moreover their powers of dispersal are great relative to those of their food and consequently they consume a large proportion of their total food supply. On the other hand, *Glossina* characteristically consumes a small proportion of the total stock of its food. It is difficult to imagine a situation in which *Glossina* is likely to experience an absolute shortage of food. Thus the proportion of the total stocks of a resource in the area that are consumed by a population experiencing a relative shortage of food may be greatly different and the degree to which the relative shortage is due to intrinsic or extrinsic causes may differ from one species to another and from one environment to another.

4. The Relationship between no Shortage, Absolute Shortage, and Relative Shortage

Figure 1 is like a map. Any persistent population of animals can be plotted on it according to its status relative to one of its resources. The ordinate represents the proportion of the population that get enough to complete their life-cycles, and the abscissa represents the abundance of the animals relative to the resource. For example *Glossina* is placed in the left-hand bottom corner (at Shinyanga one might imagine it disappearing off the map at the bottom left-hand corner); it characteristically experiences a severe relative shortage not due to its own feeding. *Cactoblastis*, as it has existed in Australia since 1940, is placed near the bottom right-hand corner; it experiences a severe relative shortage that is brought about by its own activities, as was explained in Section 3.

Most of the area of Fig. 1 represents states of relative shortage. Species which in nature experience a relative shortage are located to the right when they characteristically consume a large proportion of the resource that is available and to the left when they characteristically consume rather little of their total stocks of a resource.

Species like *Parus* and *Lucilia* are located at the extreme right-hand edge of the area of shortages; their density is usually so great relative to their resources that there are more than enough individuals in any generation to consume all the resource, nesting-sites (*Parus*) or food (*Lucilia*) in their area of distribution. Species like *Parus* and *Lucilia* that belong to category B of Table 1 can exist as extensive natural populations at densities that ensure a persistent absolute shortage.

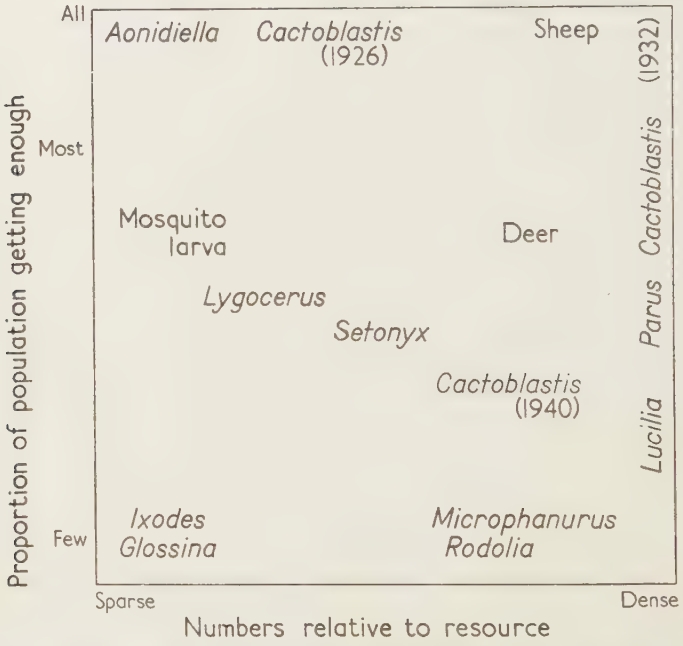


FIG. 1. The way in which resources may influence an animal's chance to survive and reproduce. For an explanation of the placing of each species on this map see text.

Absolute shortages that develop in local populations of species in category A of Table 1 such as *Cactoblastis* and *Rodolia* cannot be represented in this figure which represents only persistent populations occupying substantial areas. For example there is no place for the cats on Berlenga island; they might be thought of as sliding down the right-hand edge and disappearing off the map in the bottom right-hand corner as they become extinct.

Aonidiella in citrus groves near the coast in California and sheep on a well-managed pasture live amid an abundance of food. They are therefore placed along the extreme top edge of the figure and do not enter the area of

shortages at all. *Aonidiella* is kept rare relative to its food by *Aphytis* and is therefore placed on the left-hand side of the figure whereas the sheep, kept numerous relative to its food by the husbandry of the farmer, is placed on the right-hand side.

We have shown *Cactoblastis* on the map in three positions. In 1926 the first *Cactoblastis* in Australia found themselves amid a great abundance of prickly-pear and there were few of them. By 1932 their numbers had increased over extensive areas until there were more than enough of them to consume the supply of food. By 1940 the prickly-pear had come to be distributed sparsely and patchily and the numbers of *Cactoblastis* had declined under the influence of the resulting relative shortage of food (see above). This progression from no shortage through absolute shortage to relative shortage is indicated by the three positions of *Cactoblastis* on Fig. 1. A similar pathway on the area of shortages could be mapped for *Rodolia* and so, probably, for all species that experience a relative shortage that is due to their own activities.

We think that this "map" illustrates and summarizes all the principles that we have discussed in relation to the concept of "material resources" as a component of environment.

5. Discussion

In Andrewartha & Birch's theory one of the components of environment was called "a place in which to live". The quantitative aspect of this component concerned the supply of certain resources such as nesting-sites, living-room and so on. It is now clear that we have to think about shortages in such resources in the same way as we think about shortages in food. Dr. D. Maelzer has suggested to us that, in these circumstances, the idea of "a place in which to live" as a component of environment may be dispensed with, and we understand that he is preparing a paper on this subject which he hopes will appear shortly.

Certain sorts of animals, especially birds and rodents, show a special sort of aggressive behaviour which seems to result in keeping the animals rare relative to their resources; and the status of such animals in terms of Fig. 1 would seem to be that they constitute one rather special category of the great diversity of species that characteristically experience no shortage (Andrewartha, 1959). Wellington (1960) has recently shown that a mechanism dependent upon qualitative differences among individuals operates, though not through aggression, to keep *Malacosoma pluviale* rare relative to its food. And Chitty (1960) has suggested that some such mechanism, though not necessarily manifesting itself in either of the ways just mentioned, may be latent in all animal populations. This may be so but we think that Chitty may have underestimated the diversity of ways in which

animals may be kept rare relative to their resources. For example, in quoting Andrewartha & Birch's theory of environment Chitty considered weather, but did not mention any of the substantial body of evidence that Andrewartha & Birch presented to show that not only weather, but also any other component of environment might be important on occasion. Moreover Chitty's hypothesis does not seem to be necessary to explain any of the case-histories that we used to construct Fig. 1, nor many others that we know of.

Milne (1957, 1958, and elsewhere) has developed the theoretical argument that "intraspecific competition", being the only "perfectly density-dependent factor" is the only "factor" that can control a population. He uses "control" in the special sense of "keep extant" or "prevent from becoming extinct". When we test this idea by observing nature we find that "intra-specific competition" did not prevent the cats on the island of Berlenga from exhausting their food supply and becoming extinct, nor the mites in the cages maintained by Gause *et al.* (1936) nor the braconids in three out of six of Utida's (1957) cages. Indeed we are led to infer that on these occasions when intra-specific competition can be observed in its purity (i.e. in small uniform areas) it is likely to lead, not to "control" but to "extinction". Of course Milne would be well aware of these and similar observations, and if one reads his arguments in their full context it is clear that he takes for granted what we have called the pathway from abundance through absolute shortage to relative shortage and in his papers "intra-specific competition" implies this whole succession including the ultimate relative shortage. In other words Milne's theory seems to coincide with that part of ours which deals with this pathway. But in stating the theory verbally we would prefer a different emphasis from Milne, because it seems to us that in most circumstances intra-specific competition *per se* leads to extinction, but, in nature, species and extensive populations that might otherwise be extinguished by "intra-specific competition" are preserved by their ability to disperse and as a result enter a continuing state of relative shortage that arises from the same causes as the "intra-specific competition". The circumstances in which intra-specific competition *per se* would not lead to extinction are defined in Section B (*b*) of Table 1; in these circumstances no matter how intense the competition may be the same *number* of animals will survive and have a chance of continuing the population (Andrewartha, 1961, sec. 9.32).

Huffaker (1957), from a wide experience in the biological control of weeds, suggested that the rareness of insects relative to their food may have been overstated by Andrewartha & Birch (1954, p. 22) and the authors whom they quote, because what at first sight appears to be food may not really be so. Perhaps the animals can use only a small part of a large bush,

or a small part in a particular condition, or maybe only one rather rare genotype in a large population may be suitable for them to eat. This is an interesting idea, because if it is followed up by observation and experiment it is likely to add greatly to our understanding of relative shortages of food.

REFERENCES

Andrewartha, H. G. (1957). *Cold Spr. Harb. Symp. quant. Biol.* **22,** 219.
Andrewartha, H. G. (1959). *Aust. J. Sci.* **22,** 200.
Andrewartha, H. G. (1961). "The Ecology of Animal Populations". Methuen, London.
Andrewartha, H. G. & Birch, L. C. (1954). "The Distribution and Abundance of Animals". University of Chicago Press, Chicago.
Chitty, D. (1960). *Canad. J. Zool.* **38,** 99.
Cockerell, T. D. A. (1934). *Nature, Lond.* **133,** 329.
De Bach, P. (1949). *Ecology*, **30,** 14.
De Bach, P., Fisher, T. W. & Landi, J. (1955). *Ecology*, **36,** 743.
De Bach, P. & Sisojevic, P. (1960). *Ecology*, **41,** 153.
Dodd, A. P. (1940). "The Biological Campaign against Prickly Pear". Comm. Prickly Pear Board, Brisbane.
Elton, C. (1927). "Animal Ecology". Sidgwick and Jackson, London.
Errington, P. L. (1946). *Quart. Rev. Biol.* **21,** 145, and 241.
Gause, G. F. (1934). "The Struggle for Existence". Williams and Wilkins, Baltimore.
Gause, G. F., Smaragdova, N. P. & Witt, A. A. (1936). *J. anim. Ecol.* **5,** 1.
Huffaker, C. B. (1957). *Hilgardia*, **27,** 101.
Jackson, C. H. N. (1937). *Proc. zool. Soc. Lond.*, 1936, 811.
Kluijver, H. N. (1951). *Ardea*, **39,** 1.
Leopold, A. (1943). *Wis. Conserv. Bull.*, August 1943.
Main, A. R., Shield, J. W. & Waring, H. (1959). *Monograph. Biol.* **8,** 315.
McDougall, K. D. (1943). *Ecol. Monogr.* **13,** 323.
Milne, A. (1947). *Parasitology*, **38,** 34.
Milne, A. (1957). *Cold Spr. Harb. Symp. quant. Biol.* **22,** 253.
Milne, A. (1958). *Nature, Lond.* **182,** 1251.
Nicholson, A. J. (1947). "Fluctuation of Animal Populations", *Rep. 26th Meeting A.N.Z.A.A.S.* Perth, Western Australia.
Potts, W. H. & Jackson, C. H. N. (1953). *Bull. entom. Res.* **43,** 365.
Smith, H. S. (1939). *Ecol. Monogr.* **9,** 311.
Ullyett, G. C. (1950). *Phil. Trans.* **B234,** 77.
Utida, S. (1957). *Cold Spr. Harb. Symp. quant. Biol.* **22,** 139.
Wellington, W. G. (1960). *Canad. J. Zool.* **38,** 289.

J. Theoret. Biol. (1961) 1, 98–103

Reductions and Oxidations in Mammalian Biosyntheses

John M. Lowenstein

Graduate Department of Biochemistry, Brandeis University, Waltham, Mass., U.S.A.

(*Received* 10 *October* 1960)

An examination of the biosynthetic pathways that are now known in the mammal suggests that whenever a biosynthetic step involves a reduction, the pyridine nucleotide utilized is TPN†, and whenever a biosynthetic step involves an oxidation, the pyridine nucleotide utilized is DPN. More specifically, the reducing agent is TPNH, and the oxidizing agent is DPN. Examples of reactions which support this hypothesis are shown in Tables 1 and 2.

The reciprocal relationship in biosyntheses between TPNH and DPN is reflected by the ratios TPNH/TPN and DPNH/DPN which are found in the cell (Table 3). The ratio TPNH/TPN is usually greater than five, and the ratio DPNH/DPN is usually less than 0·7 A high ratio of TPNH/TPN is maintained in most if not all tissues even though the tissues may be respiring rapidly, whereas in the same tissues a low ratio of DPNH/DPN prevails. The two pyridine nucleotides form separate redox systems which are not in equilibrium with each other. Moreover, the total concentration of TPN plus TPNH is highest in organs that are most active biosynthetically.

The different roles of DPN and TPN have been discussed frequently. Kaplan, Swartz, Frech & Ciotti (1956) postulated that the prime function of TPNH is to serve as a reducing agent, and that of DPNH is to give rise to energy utilizable in the form of ATP. The dual functions of TPNH and DPNH have also been discussed by Horecker and Hiatt (1958), and examples of biosynthetic reductions which require TPNH have been quoted by Grant & Mongkolkul (1959), Dickens, Glock & McLean (1959) and (after this manuscript was prepared) by Klingenberg & Bücher (1960).

† The abbreviations used are: DPN and DPNH, the oxidized and reduced forms of diphosphopyridine nucleotide; TPN and TPNH, the oxidized and reduced forms of triphosphopyridine nucleotide respectively; ADP and ATP, the di- and triphosphates of adenosine; CoA, coenzyme A.

Reductive biosyntheses in the mammal which are TPNH-specific

Enzyme	Product of reaction	On pathway leading to	Remarks	References
β-Ketoacyl CoA reductase	β-hydroxyacyl CoA	fatty acids	DPNH is about one-sixth as effective as TPNH when both reductive steps are assayed together	Brady, 1958; Wakil, 1958; Formica & Brady, 1959; Wakil, Titchener & Gibson, 1959; Vagelos, 1959.
β-Hydroxyacyl CoA → fattyacyl CoA	fattyacyl CoA	fatty acids	see above	see above.
Palmityl CoA reductase	palmitaldehyde	sphingosine		Brady & Koval, 1958.
Phenylalanine hydroxylase	tyrosine	proteins	other aerobic hydroxylations are also TPNH-specific	Kaufman, 1959; Kaufman & Levenberg, 1959.
Steroid hydroxylase	(e.g.) corticosterone	various steroids	–	Hayano & Dorfman, 1954; Grant & Brownie, 1955; Leybold & Staudinger, 1959.
Fatty acid hydroxylase	hydroxy-fatty acids	unsaturated fatty acids	pyridine nucleotide specificity is not absolute, but enzyme favours TPNH over DPNH	Bloomfield & Bloch, 1960.
Glucose reductase	sorbitol	seminal fructose	in this pathway there is a net synthesis of fructose from glucose, and not vice versa.	Hers, 1956 & 1960.
Ring closure and demethylation enzymes	steroids	steroids	DPNH is much less effective than TPNH	Tchen & Bloch, 1957.
Dihydrofolate reductase	tetrahydrofolate	see remarks	this reaction is necessary to regenerate tetrahydrofolate in the methylation of deoxyuridylate to thymidylate	Futterman, 1957; Humphreys & Greenberg, 1958; Osborn & Huennekens, 1958; Friedkin, 1959; Peters & Greenberg, 1958 & 1959.
Malic enzyme	malate	hexose from pyruvate	an alternative pathway which does not involve oxido-reduction reactions may be more important quantitatively	Krebs, 1954; Rutter & Lardy, 1958; Stickland, 1959; Utter & Keech, 1960.
Glucuronate reductase	L-gulonate	L-xylulose	—	Ashwell, Kanfer & Burns, 1959.

TABLE 2

Oxidative biosyntheses in the mammal which are DPN-specific

Enzyme	Product of reaction	On pathway leading to	Remarks	References
Dihydro-orotate dehydrogenase	orotate	pyrimidine nucleotides	—	Reichard, 1959.
UDP-glucose dehydrogenase	UDP-glucuronate	polysaccharides, glucuronosides L-xylulose	—	Maxwell, Kalckar, & Strominger, 1956.
Inosinate hydroxylase	xanthidylate	guanylate	other anaerobic hyroxylations are also DPN specific	Lagerkvist, 1955; Abrams & Bentley, 1955; Gehring & Magasanik, 1955.
Sorbitol dehydrogenase	fructose	seminal fructose	in this pathway there is a net synthesis of fructose from glucose, and not vice versa	Hers, 1956 & 1960; Williams-Ashman, Banks & Wolfson, 1957.
Malate dehydrogenase	oxaloacetate	urea	oxaloacetate yields aspartate, which is used for arginine synthesis	Ochoa, 1955.
Malate dehydrogenase	oxaloacetate	hexose from pyruvate	an alternative pathway which does not involve oxido-reductions may be more important quantitatively	Utter, 1959; Krebs, 1954; Utter & Keech, 1960 and personal communication.
L-Gulonate dehydrogenase	L-xylulose	—	—	Ashwell, Kanfer & Burns, 1959.

Exceptions to the present hypothesis are the reactions catalysed by triosephosphate and α-glycerophosphate dehydrogenases. The products of these reactions, glyceraldehyde-3-phosphate and α-glycerophosphate, lie on the pathways of hexose and fat synthesis respectively. In the biosynthetic direction these reactions involve reductions which utilize DPNH and not TPNH. The equilibrium of the triosephosphate dehydrogenase reaction is heavily in favour of glyceraldehyde-3-phosphate formation (Meyerhof & Oesper, 1947), and a DPNH/DPN ratio of 0·3 does not prevent the reaction from going in the direction of glycogen synthesis. The driving force towards glycogen is not the ratio of DPNH/DPN, but the generation of 1,3-diphosphoglycerate, which in turn depends on a high ratio of ATP ADP. The equilibrium of the α-glycerophosphate dehydrogenase reaction is also heavily in favour of α-glycerophosphate.

TABLE 3

Ratios of oxidized reduced pyridine nucleotides in whole rat tissues (condensed from Glock & McLean (1955), cf. Jedeikin & Weinhouse (1955)). The difference in the magnitude of the TPNH/TPN and DPNH/DPN ratios is found for both intra- and extra-mitochondrial pyridine nucleotides (Jacobson & Kaplan, 1957)

Tissue	Coenzyme content (μg./g. tissue) DPN + DPNH	TPN + TPNH	DPNH/DPN	TPNH/TPN
Liver	574	211	0·55	34
Adrenal	469	133	0·49	7
Kidney	435	57	0·95	18
Mammary gland†	310	51	0·37	>25
Heart	483	36	0·62	8
Diaphragm	427	13	0·48	> 6
Brain	221	8	0·66	> 4

† 18 days' lactation.

The glutamate dehydrogenase reaction may or may not fit the hypothesis. The reaction occurs in both anabolic and catabolic pathways, and the enzyme is about equally active with both pyridine nucleotides. In view of the TPNH/TPN and DPNH/DPN ratios which prevail in the cell, it may be argued that the pyridine nucleotide which is used in the glutamate dehydrogenase reaction in the direction of reduction is TPNH and that which is used in the direction of oxidation is DPN. Whether this is true under physiological conditions remains to be determined. The lactate dehydrogenase reaction is difficult to fit into an anabolic or a catabolic category, and it seems reasonable to regard lactate as a hydrogen carrier rather than as a synthetic end product.

It is implicit in the above hypothesis that there must be reactions which serve to regenerate TPNH and DPN. These "regeneration reactions" need not be considered as exceptions to the hypothesis. The chief reactions which may regenerate TPNH are those catalysed by isocitrate, glucose-6-phosphate, and 6-phosphogluconate dehydrogenases. The transhydrogenase reaction, TPNH + DPN → TPN + DPNH, contributes only to the equilibration of the ratios TPNH/TPN and DPNH/DPN in the mitochondria, but experiments with isolated mitochondria suggest that the equilibration does not normally go to completion (Jacobson & Kaplan, 1957; Bücher & Klingenberg, 1958; Purvis, 1959). A simple reversal of the reaction as written cannot be responsible for the ratios TPNH/TPN and DPNH/DPN which are observed in the cell. The direct oxidation of extramitochondrial DPNH by intact mitochondria is very slow if it occurs at all (Chance & Thorell, 1959; Pullman & Racker, 1956; Lehninger, 1955). Instead the oxidation of extramitochondrial DPNH probably takes place through the intermediate agency of carrier compounds (Estabrook & Sacktor, 1958; Bücher & Klingenberg, 1958; Devlin & Bedell, 1959), but the carriers involved in mammalian metabolism have not yet been identified with certainty. Mitochondrial DPNH is oxidized by the DPNH oxidase system (Green, 1959).

This work was aided by the National Science Foundation.

REFERENCES

Abrams, R. & Bentley, M. (1955). *J. Amer. chem. Soc.* **77,** 4179.
Ashwell, G., Kanfer, J. & Burns, J. J. (1959). *J. biol. Chem.* **234,** 472.
Bloomfield, D. K. & Bloch, K. (1960). *J. biol. Chem.* **235,** 337.
Brady, R. O. (1958). *Proc. nat. Acad. Sci., Wash.* **44,** 993.
Brady, R. O. & Koval, G. J. (1958). *J. biol. Chem.* **233,** 26.
Bücher, T. & Klingenberg, M. (1958). *Angew. Chem.* **70,** 552.
Chance, B. & Thorell, B. (1959). *J. biol. Chem.* **234,** 3044.
Devlin, T. M. & Bedell, B. H. (1959). *Biochim. biophys. Acta,* **36,** 564.
Dickens, F., Glock, G. E. & McLean, P. (1959). "Ciba Foundation Symposium on the Regulation of Cell Metabolism," 150. Little, Brown & Co., Boston.
Estabrook, R. W. & Sacktor, B. (1958). *J. biol. Chem.* **233,** 1014.
Formica, J. V. & Brady, R. O. (1959). *J. Amer. chem. Soc.* **81,** 752.
Friedkin, M. (1959). *Fed. Proc.* **18,** 230.
Futterman, S. (1957). *J. biol. Chem.* **228,** 1031.
Gehring, L. B. & Magasanik, B. (1955). *J. Amer. chem. Soc.* **77,** 4685.
Glock, G. E. & McLean, P. (1955). *Biochem. J.* **61,** 388.
Grant, J. K. & Brownie, A. C. (1955). *Biochim. biophys. Acta* **18,** 433.
Grant, J. K. & Mongkolkul, K. (1959). *Biochem. J.* **71,** 34.
Green, D. E. (1959). *Advanc. Enzymol.* **21,** 73.
Hayano, M. & Dorfman, R. I. (1954). *J. biol. Chem.* **211,** 227.
Hers, H. G. (1956). *Biochim. biophys. Acta,* **22,** 202.
Hers, H. G. (1960). *Biochim. biophys. Acta,* **37,** 127.
Horecker, B. L. & Hiatt, H. H. (1958). *New Engl. J. Med.* **258,** 177 and 225.
Humphreys, G. K. & Greenberg, D. M. (1958). *Arch. Biochem. Biophys.* **78,** 275.

Jacobson, K. B. & Kaplan, N. O. (1957). *J. biol. Chem.* **226,** 603.
Jedeikin, L. A. & Weinhouse, S. (1955). *J. biol. Chem.* **213,** 271.
Kaplan, N. O., Swartz, M. N., Frech, M. E. & Ciotti, M. M. (1956). *Proc. nat. Acad. Sci., Wash.* **42,** 481.
Kaufman, S. (1959). *J. biol. Chem.* **234,** 2677.
Kaufman, S. & Levenberg, B. (1959). *J. biol. Chem.* **234,** 2683.
Klingenberg, M. & Bücher, T. (1960). *Annu. Rev. Biochem.* **29,** 669.
Krebs, H. A. (1954). *Johns Hopk. Hosp. Bull.* **95,** 19.
Lagerkvist, U. (1955). *Acta chem. scand.* **9,** 1028.
Lehninger, A. L. (1955). "The Harvey Lectures 1953-1954," 176. Academic Press, New York.
Leybold, K. & Staudinger, H. (1959). *Biochem. Z.* **331,** 389.
Maxwell, E. S., Kalckar, H. M. & Strominger, J. L. (1956). *Arch. Biochem. Biophys.* **65,** 2.
Meyerhof, O. & Oesper, P. (1947). *J. biol. Chem.* **170,** 1.
Ochoa, S. (1955). "Methods in Enzymology," Vol. I, 735. Academic Press, New York.
Osborn, M. J. & Huennekens, F. M. (1958). *J. biol. Chem.* **233,** 969.
Peters, J. M. & Greenberg, D. M. (1958). *Nature, Lond.* **181,** 1669.
Peters, J. M. & Greenberg, D. M. (1959). *Biochim. biophys. Acta,* **32,** 273.
Pullman, M. E. & Racker, E. (1956). *Science,* **123,** 1105.
Purvis, J. (1959). *Abstr. Pap. Amer. chem. Soc. 135th Meeting,* 5C.
Reichard, P. (1959). *Advanc. Enzymol.* **21,** 263.
Rutter, W. J. & Lardy, H. A. (1958). *J. biol. Chem.* **233,** 374.
Stickland, R. G. (1959). *Biochem. J.* **73,** 646, 654.
Tchen, T. T. & Bloch, K. (1957). *J. biol. Chem.* **226,** 921.
Utter, M. F. (1959). *Ann. N.Y. Acad. Sci.* **72,** 451.
Utter, M. F. & Keech, D. B. (1960). *J. biol. Chem.* **235,** PC17, and personal communication.
Vagelos, P. R. (1959). *J. Amer. chem. Soc.* **81,** 4119.
Wakil, S. J. (1958). *J. Amer. chem. Soc.* **80,** 6465.
Wakil, S. J., Titchener, E. B. & Gibson, D. M. (1959). *Biochim. biophys. Acta,* **34,** 227.
Williams-Ashman, H. G., Banks, J. & Wolfson, S. K. (1957). *Arch. Biochem. Biophys.* **72,** 485.

J. Theoret. Biol. (1961) **1**, 104–106

Surface Extension as the Mechanism of Cellular Movement and Cell Division

L. G. E. Bell

Department of Zoology, King's College, London, England

(*Received* 12 *October* 1960)

The main process occurring during cellular movement, as seen in amoebae and migrating cells in tissue culture, is the formation of new surface. Localized surface formation gives rise to directional movement such as the formation of pseudopods.

It is suggested that new surface is formed by the local production of surface material. The simple proteins of cells do not seem likely to be this surface material, as local production would not necessarily put more protein into the surface; indeed, the continued production of protein would have the effect of blowing the cell up like a balloon with no surface extension until the spherical form was attained. Among the cell materials which have the required property for a surface material of strong adsorption to protein are the complex polysaccharides occurring as mucoproteins. The sequence of events suggested is that local production of these polysaccharides occurs and that their physical properties then cause them to be adsorbed at the protein/external medium interface. Because there is already a close packing of charged polysaccharide groupings in the surface, more surface has to be made to accommodate the new surface material. The production of the new material and its adsorption constitutes the work done in cellular movement. Polarity of movement would result from the disappearance of surface material at other sites.

Cytoplasmic currents and sol-gel changes are, in this scheme, relegated to a class of secondary phenomena.

The results of the periodate oxidation and other techniques for the histochemical demonstration of mucoproteins, and the results of immunochemistry, show that many cellular surfaces consist largely of a mucoprotein, i.e. polysaccharide, component. This is certainly so in amoebae, and recent electron micrographs (Mercer, 1959; Pappas, 1959) show this layer to consist of closely packed elongated structures. It is also known, particularly from amoebae (Marshall, Schumaker & Brandt, 1959; Holter, 1959), that this material returns to the interior of the cell during pinocytosis, and it is possible that a similar mechanism might account for

the local absorption of surface material during locomotion, or at least that a basic mechanism is common to both phenomena.

In cell division the prime event is an increase in cell surface, and here, too, surface polysaccharide is involved. It has been well known for many years that a plate of polysaccharide material is laid down as the first stage in cell division (as distinct from nuclear division) in plant cells. It is known from the work of Selman & Waddington (1955) that a discrete layer is formed in animal cells ahead of the cleavage furrow and it seems very likely, on the basis of the hypothesis proposed here, that this also involves a polysaccharide-containing layer. Its nature could possibly be elucidated by the use of fluorescent antibody techniques, using labelled antibody to the surface antigens. The mechanism which would fit well with the often described surface movements of the dividing egg would suppose that the layer formed ahead of the cleavage furrow was due to the taking in of surface material by a mechanism similar to that operating during pinocytosis and that new surface material was produced at the poles. This could also account for the appearance of new surface in the interior of dividing plant cells. The slow rate of cleavage in yolky cells is caused by the obstruction offered to the aggregation of surface material taken into the interior. The yolk granules have to be displaced and their position occupied by surface material. If mechanisms dependent on a contractile cortical zone are the cause of cell cleavage (Wolpert, 1960), then only friction between yolk granules could explain the slow rate; this seems unlikely.

Production of surface polysaccharide could involve either synthesis, or release from a previous complex. That there may be a structural relationship between the surface-forming material and another component is indicated by the explosive bubbling of the surface which may accompany cell division. This sudden surface extension seems likely to be due to the release of preformed surface material rather than very rapid synthesis.

The actions of ribonuclease on amoebae and other cells (Brachet, 1957) suggest that an association of ribonucleic acid and surface material should be looked for, whether or not polysaccharide is the other component.

The idea that surface extension by production of surface material is the basis of much cellular activity can be applied to differentiation and morphogenesis with its associated movements. Contact between cells would give local inhibition of surface expansion if chemical groupings in one surface interacted with groupings in the other surface, so as to lock the surfaces together. Clearly, groupings in the surface of the same cell can interact similarly with each other to give regional differentiation in the surface properties of the cell. This indicates that the basis of much cellular differentiation could be local production of surface material with different arrangements of reactive chemical groups which would react

among themselves or with surface groups of other cells to modify locally the surface extensibility and certainly then the shape of the cell. That there are many such specific groupings is indicated by immuno-chemistry. This suggestion shows how such cell-specific proteins may in one way take part in morphological differentiation. Competence to react to an external differentiating agent would involve a cell having surface material whose reactive groups could respond to the agent by combining with it. This is a mechanism suggested by Weiss, but here it is referred to an increase or decrease of surface extensibility in the reacting region.

The morphological changes seen when cells are first transferred to an alien medium for tissue culture are explicable if the state of affairs is as proposed here.

The continued cell division and the locomotory properties of malignant cells might both be due to a derangement of the surface-producing mechanism.

REFERENCES

Brachet, J. (1957). "Biochemical Cytology". Academic Press, New York.
Holter, H. (1959). *Ann. N.Y. Acad. Sci.* **78,** (2), 524.
Marshall, J. M., Schumaker, V. N. & Brandt, P. W. (1959). *Ann. N.Y. Acad. Sci.* **78,** (2), 515.
Mercer, E. H. (1959). *Proc. roy. Soc.* **B150,** 216.
Pappas, G. D. (1959). *Ann. N.Y. Acad. Sci.* **78,** (2), 448.
Selman, G. G. & Waddington, C. H. (1955). *J. exp. Biol.* **32,** 700.
Weiss, P. (1939). "The Principles of Development". Holt, New York.
Wolpert, L. (1960). *Int. Rev. Cytol.* **10,** 163.

Instructions to Authors

1. The *Journal of Theoretical Biology* is intended to be a central publishing medium for theoretical work in all fields of biology.

Appropriate subjects for papers will include:

(*a*) generalized theories;
(*b*) theories of specific processes or phenomena, e.g. crossing over, learning, active transport, photosynthesis, carcinogenesis;
(*c*) theoretical discussion of specific projects, e.g. selective mutation, drug design, electronic models;
(*d*) theoretical discussion of methods, e.g. cytochemical methods and instrumentation, amplification techniques;

Papers concerned mainly with experimental work will not normally be accepted, but a paper will not be rejected simply because it contains some new experimental material.

2. Reviews will be accepted only if they include a substantial new theoretical discussion. Letters to the editor, intended to direct attention to a particular theoretical point, will be accepted.

3. *Submission of Manuscripts.* Manuscripts should be sent to Professor J. F. Danielli, Zoology Department, King's College, London, or to any member of the Editorial Board. The submission of manuscript will be taken to imply that the material is original.

Papers may be submitted in English, French or German. It is emphasized however that papers in English will reach a much larger group of readers than those in other languages. A paper submitted in French or German should have an abstract in English.

Publication will be accelerated if two copies of a manuscript are submitted.

4. *Manuscripts.* If the Journal is to serve its intended purpose, it is essential that the main body of its contents should be intelligible to all biologists. To this end, wherever possible mathematical development should be confined to appendices, and specialized terms should be clearly defined in the introduction to a paper.

Summary. Every paper should be preceded by a summary: this should not normally exceed 300 words, and should be intelligible to the general reader without reference to the main text. Abbreviations should not be used.

Lay-out. Manuscripts should be typewritten in double spacing throughout (including summary, footnotes, tables and legends) on one side of quarto paper, with wide margins. Footnotes, tables and legends for illustrations should be typed separately at the end of the manuscript.

Running Headline. A short running title of not more than 40 characters (including spaces), suitable for page headings, should be given if the full title is longer than this.

5. *Conventions.* The conventions used in current issues of the Journal for headings, references, etc., should be followed in preparing manuscripts. References should be listed alphabetically at the end of the paper, and abbreviated according to the *World List of Scientific Periodicals.* Citations in the text should read thus: Smith, Brown & Robinson (1957), or (Smith, Brown & Robinson, 1957). When a citation has more than two authors, the names of all of them should be given at the first citation (except for more than five authors), and on subsequent citations thus:

(Smith *et al.*, 1957). The conventions (Brown, 1957a), (Brown, 1957b) should be used when more than one paper by the same author(s) has appeared in one year. Volume number should be indicated in the typescript thus: 53.

Footnotes should be avoided whenever possible. Customary abbreviations will be used; authors may refer to *Chemical Abstracts*, to the *Biochemical Journal*, **66**, 8 (1957), and to *Notes on the preparation of papers to be communicated to the Royal Society* (1951). Certain non-standard abbreviations of chemical substances are convenient, and those listed in the *Journal of Biological Chemistry* may be used without further definition (e.g. DNA, ATP). Any other abbreviations of this kind should be used sparingly, and must be defined in a footnote at the point in each manuscript where the first such abbreviation occurs.

6. *Illustrations.* It is often convenient to make a drawing two or three times larger than it will appear in the Journal, but very large drawings should be avoided and it will be helpful if drawings and illustrations are submitted on quarto sheets. Drawings should be made in Indian ink on tracing linen, Bristol board, or faintly blue-lined graph paper; the originals should always be submitted with the manuscript, since photo-prints of drawings cannot be reproduced satisfactorily.

Line drawings will be referred to as Fig. 1, Fig. 2, etc., and photographs as Plate I, Plate II, etc. The name(s) of the author(s) and the number should be written on all illustrations, together with the indication "Top" if necessary. Besides the original illustrations *one* duplicate set must accompany the second copy of the manuscript; these may be photo-prints, and need not be of the same quality as the originals but must be clear and legible. If oversize original drawings are submitted (bigger than 8 by 10 inches) they must be accompanied by *two* sets of small photo-prints.

7. *Reprints.* Fifty reprints of each paper will be supplied free of charge; additional copies may be purchased. Order forms will be sent with proofs.

Contents

Printed in Great Britain by The Whitefriars Press Ltd., London and Tonbridge